

KW	termination sequence.	PR	28-JUN-1999;	99US-0140823.
XX		PR	29-JUN-1999;	99US-0140991.
OS	Arabidopsis thaliana.	PR	30-JUN-1999;	99US-0141287.
XX		PR	01-JUL-1999;	99US-0141842.
PN	EP1033405-A2.	PR	01-JUL-1999;	99US-0142154.
XX		PR	02-JUL-1999;	99US-0142055.
PD	06-SEP-2000.	PR	06-JUL-1999;	99US-0142390.
XX		PR	08-JUL-1999;	99US-0142803.
XX		PR	09-JUL-1999;	99US-0142920.
XX		PR	12-JUL-1999;	99US-0142977.
XX		PR	13-JUL-1999;	99US-0143542.
PR	25-FEB-1999;	PR	14-JUL-1999;	99US-0143624.
PR	05-MAR-1999;	PR	15-JUL-1999;	99US-0144005.
PR	09-MAR-1999;	PR	16-JUL-1999;	99US-0144085.
PR	23-MAR-1999;	PR	16-JUL-1999;	99US-0144086.
PR	25-MAR-1999;	PR	19-JUL-1999;	99US-0144325.
PR	29-MAR-1999;	PR	19-JUL-1999;	99US-0144331.
PR	01-APR-1999;	PR	19-JUL-1999;	99US-0144332.
PR	06-APR-1999;	PR	19-JUL-1999;	99US-0144333.
PR	08-APR-1999;	PR	19-JUL-1999;	99US-0144334.
PR	16-APR-1999;	PR	19-JUL-1999;	99US-0144335.
PR	19-APR-1999;	PR	20-JUL-1999;	99US-0144352.
PR	21-APR-1999;	PR	20-JUL-1999;	99US-0144632.
PR	23-APR-1999;	PR	20-JUL-1999;	99US-0144884.
PR	28-APR-1999;	PR	21-JUL-1999;	99US-0144814.
PR	30-APR-1999;	PR	21-JUL-1999;	99US-0145086.
PR	30-APR-1999;	PR	21-JUL-1999;	99US-0145088.
PR	30-APR-1999;	PR	22-JUL-1999;	99US-0145085.
PR	04-MAY-1999;	PR	22-JUL-1999;	99US-0145087.
PR	05-MAY-1999;	PR	22-JUL-1999;	99US-0145089.
PR	06-MAY-1999;	PR	23-JUL-1999;	99US-0145192.
PR	06-MAY-1999;	PR	23-JUL-1999;	99US-0145145.
PR	07-MAY-1999;	PR	23-JUL-1999;	99US-0145218.
PR	11-MAY-1999;	PR	23-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	PR	26-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	PR	26-JUL-1999;	99US-0145913.
PR	14-MAY-1999;	PR	27-JUL-1999;	99US-0145918.
PR	14-MAY-1999;	PR	27-JUL-1999;	99US-0145919.
PR	18-MAY-1999;	PR	28-JUL-1999;	99US-0145951.
PR	19-MAY-1999;	PR	28-JUL-1999;	99US-0146386.
PR	20-MAY-1999;	PR	02-AUG-1999;	99US-0146388.
PR	21-MAY-1999;	PR	02-AUG-1999;	99US-0146389.
PR	24-MAY-1999;	PR	02-AUG-1999;	99US-0147038.
PR	25-MAY-1999;	PR	03-AUG-1999;	99US-0147204.
PR	27-MAY-1999;	PR	04-AUG-1999;	99US-0147302.
PR	28-MAY-1999;	PR	05-AUG-1999;	99US-0147192.
PR	01-JUN-1999;	PR	05-AUG-1999;	99US-0147260.
PR	03-JUN-1999;	PR	05-AUG-1999;	99US-0147303.
PR	04-JUN-1999;	PR	06-AUG-1999;	99US-0147416.
PR	07-JUN-1999;	PR	06-AUG-1999;	99US-0147493.
PR	08-JUN-1999;	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	PR	09-AUG-1999;	99US-0148171.
PR	14-JUN-1999;	PR	10-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	PR	11-AUG-1999;	99US-0148341.
PR	16-JUN-1999;	PR	12-AUG-1999;	99US-0148565.
PR	17-JUN-1999;	PR	13-AUG-1999;	99US-0148684.
PR	17-JUN-1999;	PR	13-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	PR	16-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	PR	17-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	PR	23-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	PR	25-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	PR	27-AUG-1999;	99US-0151066.
PR	18-JUN-1999;	PR	27-AUG-1999;	99US-0151080.
PR	21-JUN-1999;	PR	27-AUG-1999;	99US-0151303.
PR	22-JUN-1999;	PR	30-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	PR	01-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	PR	07-SEP-1999;	99US-0152363.



PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139464.  
PR 18-JUN-1999; 99US-0139465.  
PR 18-JUN-1999; 99US-0139466.  
PR 18-JUN-1999; 99US-0139467.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 28-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 22-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161921.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 9.6%; Score 296.5; DB 21; Length 532;  
Best Local Similarity 22.7%; Pred. No. 1.2e-17;  
Matches 128; Conservative 96; Mismatches 235; Indels 105; Gaps 19;

[illegible]

PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144332.  
PR 20-JUL-1999; 99US-0144337.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0144586.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 16-AUG-1999; 99US-0148684.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 23-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 25-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151085.  
PR 27-AUG-1999; 99US-0151086.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.

PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159684.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
PR 29-OCT-1999; 99US-0162143.  
PR 29-OCT-1999; 99US-0162144.

Query Match 9.6%; Score 296.5; DB 21; Length 532;  
Best Local Similarity 22.7%; Pred. No. 1.2e-17; Indels 105; Gaps 19;  
Matches 128; Conservative 96; Mismatches 235;

Qy 59 IDELAAPSVAVETFRQ-----FDDLOKHIESIENAIKLENGVVLARNNF 108  
Db 10 LTELAA-QPSFBEFQKTSILMTSCILLMOELSDHFTSLQNLKKEALKQMIETLDNQT 67  
Qy 109 HQPMLSPPRNNNSVETTV-----SQPSQEIPTVETSNKEGEG----- 146  
Db 68 QTSLESLKREVTIDHSVEIVAKVGERARAALEKARDCGGSDGDDVDEGL 127  
Qy 147 -----RMCELMCSGLRKVIYANISDAQKMBEIPSAKLAKPAKFVLDICIGFYLOQR 201  
Db 128 LSALKSLCLQMDARGFWNFVARKKELENLRSKI PAALVDCVDPAMLVLEAISEVFPVDT 187  
Qy 202 RAFTKESPMSSARQVSLILLESFLMPDRCKGVKIESWIKDEAETA AVANPKRLMTGG 261  
Db 188 RGDKVNDYGNACWVILESLTPVIVDPVIGKSRLLVTPSVKEKAKEIAETWKKSLEERGR 247  
Qy 262 LAAAEKMDARGLLLVACFGVPSNFRSTDLDLIRMGSGNEIAGALKRSQF-----L 313  
Db 248 IENVKTPDVHTFLOHLVTFGI---VKSEDLALYKLI-----VVGSAWRKQMPKLAVSVGL 299  
Qy 314 VPMVSGIVESSIKRGMHIEALEMVTYFGMEDKPSAALVLTSLKMKSKESFER-AKRKAQS 372  
Db 300 GDQMPDMIEELISRGQQLDAVHFTVEVGLVDFPFPVPLLKAYLRDAKKAASIMEDSNT 359  
Qy 373 PLAFKEAATKQALVLSVMQCMETHKLDPAKLPQWQKEQIVSLEKDTLQLDKEMEKA 432  
Db 360 GRATHLVARKEQSAKAVLKACIEEYKLE--EFFPENLKKRLDQLEKTEKKEKPAAPA 417  
Qy 433 RSLSLMBEERAAKRMVNOQIKRPLS---PMEMPPVTSSSYPIYDRSFFSQRDDQDE 489  
Db 418 N-----KXTRASYNGF--MPPAKAGRITNAYVS-SFPFIRSPSHSP 455  
Qy 490 ISALVSSYLGPSTSFPHRSRSPV-----MVLPHGGLGRSVAYEHL-----APN 536  
Db 456 QYASPAIYSPPTTV--YNNRSPFPYSPYSPYSPYSPYSPYSPYSPYSPYSPYSPYSPY 513  
Qy 537 SYSPGHGHLRHQYSPSLVHGQRH 560  
Db 514 VYHP-HHHQ-HHQF-----HHQOH 530

RESULT 10  
AAG11240

ID AAG11240 standard; Protein; 505 AA.  
XX AAG11240;  
AC AAG11240;  
DT 17-OCT-2000 (first entry)  
XX XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 9883.  
XX XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX XX  
OS Arabidopsis thaliana.  
XX XX  
PN EP1033405-A2.  
XX XX  
PD 06-SEP-2000.  
XX XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0128845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149923.  
PR 23-AUG-1999; 99US-0149902.



PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145919.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149388.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157153.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 14-OCT-1999; 99US-0159296.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 23-OCT-1999; 99US-0161406.  
PR 25-OCT-1999; 99US-0161406.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.

PR	23-MAR-1999	99US-0125780
PR	24-MAR-1999	99US-0125781
PR	25-MAR-1999	99US-0125864
PR	26-MAR-1999	99US-0125875
PR	01-APR-1999	99US-0127462
PR	06-APR-1999	99US-0128234
PR	08-APR-1999	99US-0128714
PR	16-APR-1999	99US-0129845
PR	19-APR-1999	99US-0130077
PR	21-APR-1999	99US-0130449
PR	23-APR-1999	99US-0130510
PR	23-APR-1999	99US-0130691
PR	28-APR-1999	99US-0131449
PR	30-APR-1999	99US-0132048
PR	30-APR-1999	99US-0132047
PR	04-MAY-1999	99US-0132484
PR	05-MAY-1999	99US-0132484
PR	06-MAY-1999	99US-0132486
PR	06-MAY-1999	99US-0132487
PR	07-MAY-1999	99US-0132863
PR	11-MAY-1999	99US-0134256
PR	14-MAY-1999	99US-0134218
PR	14-MAY-1999	99US-0134219
PR	14-MAY-1999	99US-0134221
PR	14-MAY-1999	99US-0134370
PR	18-MAY-1999	99US-0134376
PR	19-MAY-1999	99US-0134941
PR	20-MAY-1999	99US-0135124
PR	21-MAY-1999	99US-0135353
PR	24-MAY-1999	99US-0135629
PR	25-MAY-1999	99US-0136021
PR	27-MAY-1999	99US-0136392
PR	28-MAY-1999	99US-0136782
PR	01-JUN-1999	99US-0137222
PR	03-JUN-1999	99US-0137528
PR	04-JUN-1999	99US-0137502
PR	07-JUN-1999	99US-0137724
PR	08-JUN-1999	99US-0138094
PR	10-JUN-1999	99US-0138540
PR	10-JUN-1999	99US-0138847
PR	14-JUN-1999	99US-0139119
PR	16-JUN-1999	99US-0139452
PR	16-JUN-1999	99US-0139453
PR	17-JUN-1999	99US-0139492
PR	18-JUN-1999	99US-0139454
PR	18-JUN-1999	99US-0139455
PR	18-JUN-1999	99US-0139456
PR	18-JUN-1999	99US-0139750
PR	18-JUN-1999	99US-0139763
PR	21-JUN-1999	99US-0139817
PR	22-JUN-1999	99US-0139899
PR	23-JUN-1999	99US-0140353
PR	23-JUN-1999	99US-0140354
PR	24-JUN-1999	99US-0146695
PR	28-JUN-1999	99US-0146823
PR	29-JUN-1999	99US-0146991
PR	30-JUN-1999	99US-0141287
PR	01-JUL-1999	99US-0141842
PR	01-JUL-1999	99US-0142154
PR	02-JUL-1999	99US-0142055
PR	06-JUL-1999	99US-0142390
PR	08-JUL-1999	99US-0142803
PR	09-JUL-1999	99US-0142920
PR	12-JUL-1999	99US-0142977
PR	13-JUL-1999	99US-0143542
PR	14-JUL-1999	99US-0143624
PR	15-JUL-1999	99US-0144005

PR	16-JUL-1999;	99US-0144085.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144325.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144884.	PR	14-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144814.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	26-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160989.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161359.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146388.	PR	28-OCT-1999;	99US-0161920.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161922.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147204.	PR	29-OCT-1999;	99US-0162142.
PR	04-AUG-1999;	99US-0147302.			
PR	05-AUG-1999;	99US-0147192.			
PR	05-AUG-1999;	99US-0147260.			
PR	06-AUG-1999;	99US-0147303.			
PR	06-AUG-1999;	99US-0147416.			
PR	09-AUG-1999;	99US-0147493.			
PR	09-AUG-1999;	99US-0147935.			
PR	10-AUG-1999;	99US-0148171.			
PR	11-AUG-1999;	99US-0148319.			
PR	12-AUG-1999;	99US-0148341.			
PR	13-AUG-1999;	99US-0148565.			
PR	13-AUG-1999;	99US-0148684.			
PR	16-AUG-1999;	99US-0149368.			
PR	17-AUG-1999;	99US-0149175.			
PR	18-AUG-1999;	99US-0149426.			
PR	20-AUG-1999;	99US-0149722.			
PR	20-AUG-1999;	99US-0149723.			
PR	20-AUG-1999;	99US-0149929.			
PR	23-AUG-1999;	99US-0149902.			
PR	23-AUG-1999;	99US-0149930.			
PR	25-AUG-1999;	99US-0150566.			
PR	26-AUG-1999;	99US-0150884.			
PR	27-AUG-1999;	99US-0151065.			
PR	27-AUG-1999;	99US-0151066.			
PR	27-AUG-1999;	99US-0151080.			
PR	30-AUG-1999;	99US-0151303.			
PR	31-AUG-1999;	99US-0151438.			
PR	01-SEP-1999;	99US-0151930.			
PR	07-SEP-1999;	99US-0152363.			
PR	10-SEP-1999;	99US-0153070.			
PR	13-SEP-1999;	99US-0153758.			
PR	15-SEP-1999;	99US-0154018.			
PR	16-SEP-1999;	99US-0154039.			
PR	20-SEP-1999;	99US-0154779.			
PR	22-SEP-1999;	99US-0155139.			
PR	23-SEP-1999;	99US-0155486.			
PR	24-SEP-1999;	99US-0155659.			
PR	28-SEP-1999;	99US-0156458.			
PR	29-SEP-1999;	99US-0156596.			
PR	04-OCT-1999;	99US-0157117.			
PR	05-OCT-1999;	99US-0157753.			
PR	06-OCT-1999;	99US-0157865.			
Query Match 9.1%; Score 282; DB 21; Length 532;			Best Local Similarity 20.8%; Pred. No. 2.5e-16;		
Matches 126; Conservative 98; Mismatches 224; Indels 158; Gaps 21;					
QY	49	OSKOPOFLKSIDELAAFSVAVETPKRO-----FDDLOKHIESIENAIKSKLESNG 98	QY	99	VLAARNNNHQPMLSPRRNNVSVEITTVVS-----OPSOEIVPETSNDP 143
DB	12	KSSQSPFE-----FOKQASLMTSCNLLWKELSEHFTSMQNLMMKKSEALR 57	DB	58	QMIETLDNQTOSTIELLKREVTIDHSVEIAEGKVEERAAALDSLEKARDCGDEPTGEV 117
QY	144	EGG-----RMCELMCSKGLRKYIYANISDOAKLMEIIPSAKLAKPAKVLDCIGK 195	QY	118	DDGGLLSALKSLCKMDARGFWGFIARKKELENRSQIPVALVDCVPPKLVLEAVSE 177
DB	196	FYLOGRRAFTKESPMSSARQVSLILLESF---LMPDRGKGKYKIESWIKDEABTAAVAW 252	QY	178	VFPVDKRG-GCEKVSNDFGWACVVILESLIPVMVDPVMGKSRLLVTPSVKERAKETAETW 236
QY	253	RKRLMTEGLAAAEKMDARGILLLVACFGVPSNFRSTDLILLIRMSGNEIAGALKRSQP 312	DB	237	KASLEERGGENVKTPTDHTFLOHLVTFGIVKK-----DDLALYR-----KLIVGSAWRKOM 288
DB	313	-----LVPWVGIVESSTIKGMHIEALEMVYTFGMDKFSAAALVTSFLKMSK----- 360	QY	289	PKLAVSVGLGQMPDMEIELIRGQQLDAVHFTFEVLGVLHFPVPLLKAYLRDAKKATA 348
QY	361	---ESFERAKKAQSPKAPKEATKQLAVLSVMQCMETHKLDPAKELPGWQIKEOIVSL 417	DB	349	LITDDSNNSGRSAHL-----VAKEQSALRAVLKICEYKLE--EEFPPENLKGLDQL 400
QY	418	EKDTQLDLDKEMEERKARSLMEBAALAKRMVNOQIKRPLRS-EMEMPPTVTSSTSSYDIYRD 476	DB	401	EKTTEKEKPAVIPAN-----KXTRASYSGPMPPAKAGHIT----- 436
QY	477	RSFPQRDDQDEISALVSYLGPSTSFPHRSRRSPBYMVPLPHGLGLGRSVAYEHLAPN 536	DB	437	-----NAYVSSFPPTTFFRSQHSFQYGVF-----AYTTSPTT 471
QY	537	SY---SPGHGHLRHROQVSPSLVHG--QRHPLQY-----SPPI-HGQQOLPY 576			

```
Db 472 IYSNRSPY-----QISPEAVHGSYQTSVSYPTAYGTCSPVAAPPVYHPHPHHH 525
QY 577 GIORVY 582
Db 526 HIOHAY 531

RESULT 13
AAG11241
ID AAG11241 standard; Protein; 483 AA.
AC AAG11241;
XX
XX 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 9884.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
OS
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127482.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132407.
XX 05-MAY-1999; 99US-0132484.
XX 06-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 07-MAY-1999; 99US-0132487.
XX 11-MAY-1999; 99US-0132863.
XX 14-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134219.
XX 18-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
```









Search completed: August 14, 2003, 10:37:33  
Job time : 88 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 14, 2003, 10:36:04 ; Search time 29 Seconds  
(without alignments)  
888,528 Million cell updates/sec

Title: US-09-890-475-1

Perfect score: 3104

Sequence: 1 MSNYPPTVAQPTTTANPL.....RYLGLSNQRSPRSNSLDPK 609

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/iaa/5A COMB.pep:\*  
2: /cgn2\_6/prodata/1/iaa/5B COMB.pep:\*  
3: /cgn2\_6/prodata/1/iaa/5A COMB.pep:\*  
4: /cgn2\_6/prodata/1/iaa/5B COMB.pep:\*  
5: /cgn2\_6/prodata/1/iaa/5A COMB.pep:\*  
6: /cgn2\_6/prodata/1/iaa/5B COMB.pep:\*  
7: /cgn2\_6/prodata/1/iaa/5A COMB.pep:\*  
8: /cgn2\_6/prodata/1/iaa/5B COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match %	Score	Length	ID	Description
1	143.5	4.6	1201	3	US-09-098-901-2 Sequence 2, Appli
2	137.5	4.4	2074	4	US-09-491-356C-9 Sequence 9, Appli
3	123.5	4.0	2023	4	US-09-491-356C-8 Sequence 8, Appli
4	122.5	3.9	2285	4	US-09-308-375-2 Sequence 2, Appli
5	120.5	3.9	1001	3	US-09-060-410-2 Sequence 2, Appli
6	120.5	3.9	1001	4	US-09-723-458-2 Sequence 2, Appli
7	119.5	3.8	1388	2	US-08-585-576-1 Sequence 1, Appli
8	117.5	3.8	828	1	US-08-361-304-2 Sequence 2, Appli
9	113.5	3.7	816	2	US-08-533-306A-6 Sequence 6, Appli
10	113.5	3.7	816	2	US-08-742-922A-6 Sequence 6, Appli
11	113.5	3.7	885	2	US-08-533-306A-4 Sequence 4, Appli
12	113.5	3.7	885	2	US-08-742-922A-4 Sequence 4, Appli
13	112	3.6	3248	1	US-08-353-700-1 Sequence 1, Appli
14	112	3.6	3248	5	PCT-US95-16216-1 Sequence 6, Appli
15	110	3.5	2482	1	US-08-328-254-6 Sequence 1, Appli
16	109	3.5	578	3	US-09-066-046-6 Sequence 6, Appli
17	109	3.5	578	3	US-08-975-762-50 Sequence 50, Appl
18	109	3.5	578	3	US-09-295-028-50 Sequence 50, Appl
19	109	3.5	578	4	US-09-106-582-50 Sequence 36, Appl
20	109	3.5	1174	2	US-08-446-345-36 Sequence 36, Appl
21	108.5	3.5	1093	4	US-09-252-991A-28407 Sequence 28407, A
22	108.5	3.5	1093	5	PCT-US93-03077-1 Sequence 1, Appli
23	107.5	3.5	736	4	US-09-252-991A-21498 Sequence 21498, A
24	106	3.4	1618	1	US-07-853-913-4 Sequence 4, Appli
25	106	3.4	2293	3	US-09-368-590-2 Sequence 2, Appli
26	105	3.4	935	4	US-09-914-259-25 Sequence 25, Appl
27	104.5	3.4	784	4	US-09-740-235-1 Sequence 1, Appli

28	104.5	3.4	1075	2	US-08-993-228-19 Sequence 19, Appli
29	104.5	3.4	1196	1	US-08-144-121-4 Sequence 4, Appli
30	104.5	3.4	1196	2	US-08-735-893-4 Sequence 4, Appli
31	104.5	3.4	1786	3	US-08-973-462-8 Sequence 8, Appli
32	104	3.4	1886	4	US-08-938-105-3 Sequence 3, Appli
33	104	3.4	1839	4	US-09-310-187A-1 Sequence 1, Appli
34	103.5	3.3	749	1	US-08-369-796-8 Sequence 8, Appli
35	103.5	3.3	749	2	US-08-852-091-8 Sequence 8, Appli
36	103.5	3.3	749	3	US-08-820-754-8 Sequence 8, Appli
37	103.5	3.3	749	3	US-08-956-652-8 Sequence 8, Appli
38	103.5	3.3	749	3	US-08-956-869-8 Sequence 8, Appli
39	103.5	3.3	749	3	US-09-012-710-7 Sequence 7, Appli
40	103.5	3.3	749	3	US-08-948-547-8 Sequence 8, Appli
41	103.5	3.3	749	4	US-09-556-273-7 Sequence 7, Appli
42	103.5	3.3	749	4	US-08-956-653A-8 Sequence 8, Appli
43	103.5	3.3	749	5	PCT-US95-17025-8 Sequence 8, Appli
44	102.5	3.3	737	1	US-08-188-582-16 Sequence 16, Appl
45	102.5	3.3	737	1	US-08-646-715-16 Sequence 16, Appl

ALIGNMENTS

RESULT 1

US-09-098-901-2  
; Sequence 2, Application US/09098901B  
; Patent No. 6218144  
; GENERAL INFORMATION:  
; APPLICANT: Scott, Matthew  
; TITLE OF INVENTION: Costal2 Genes and their Uses  
; FILE REFERENCE: SUN-65P  
; CURRENT APPLICATION NUMBER: US/09/098,901B  
; CURRENT FILING DATE: 1998-06-17  
; EARLIER APPLICATION NUMBER: 60/051,347  
; EARLIER FILING DATE: 1997-06-30  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1201  
; TYPE: PRT  
; ORGANISM: D. Melanogaster  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(1201)  
; OTHER INFORMATION: Xaa = Any Amino Acid  
US-09-098-901-2

Query Match 4.6%; Score 143.5; DB 3; Length 1201;  
Best Local Similarity 19.9%; Pred. No. 0.00011;  
Matches 133; Conservative 100; Mismatches 232; Indels 203; Gaps 30;

Qy	18	PLIQRHQSQRRELPKIVETESTMDITIGSKQOPFLKSIDELAAPSVAVETFKQFD	77
Db	498	FILEABEPEDVNSEANSESPNSDNDTDSHRPDLDKIESL-----MEBFDRKTD	551
Qy	78	D-----LOKHIESIENAIKSLKES-----NG-----VVLARNNNFHOMLSPPRNN	119
Db	552	ALLLEKHAELVSKPKAVMQSDRIEIAQPEENGDDKVKVIGSRRSV-OPGASLSAE	610
Qy	120	VSVEVTTVVSQSQSIVPET-----SNKPEGRCMELMCS-----KGLRKYIYA	163
Db	611	LAMLNVAQQPFPIDPESVVDPLESSGEGIRQAALAAATAPIBQLQKLRKLV-A	669
Qy	164	NISDOAKLMEEIPSAKLA-----KEPAKFLVDC-----IGK	195
Db	670	EIEGKQRQLREIEEIQVKNNIAELVKNSTRSHAKQRFHKRAKLEAECDKAKQLGK	729
Qy	196	FYVQGR-----RAFTKESPM-----SSARQVSL-----ILSFILMPDRG	231
Db	730	ALVQGRGQSEIERWTITIGHLERRLEDLSMKHAGESCQVKVKKLQSQVGSERKQADLQ	789
Qy	232	KGKVKIESIKDEAEATAVAKRLMTSGGLAAE---KMDARGLLLVACFGVPSNFRS	288

```
Db 790 K-KRKECKLRQMEAEVLKRESRETKELVKQGGPEQGRQKAVQA-----RI 840
QY 289 TDLILLRMSGN-ELAGALKRSQPL--VPMVSGIVESSIKRGWH-----330
Db 841 THLNHILREKSNLXQPGPEQOETLRHEIRNLGTRDLLKCHDLKRLKRDVLTQK 900
QY 331 -----TEALEMVTYTFGMEKFSAAVLTSFLKMSXESPERAKRKAQSPKAPKEA 379
Db 901 BERKLECEAEIAADAEFPNEMITCHRSIDTSD-RIQREKEQOMLARNLRLSTEEM 959
QY 380 ATKQLAVLSSVMQC-----METHKLDPAKELPGWQKEQIVSL-----EKDTLQL 424
Db 960 RTLLYKYFTKVDLRDSRKLEQLVQLEBERDAWEMKERVLSNAVQAQALEGNAVL 1019
QY 425 DKEMBEKARSLSM-----EBAALAKRMYNQOIKRPLSPMEMPPVTSSS---YSPITYRD 476
Db 1020 QRCHENK---LTLMLRHAETSSASSASYGERALAP---ACVAPVQASDDFYDHYFKG 1073
QY 477 RSFPQQRDDQDEISALVSSYLGPTSTPHRSRSPSPVMDPLPHGGL-----GR 525
Db 1074 GGNPSK-----ALTI-----KAPK---PMPTGSALDKYKQKORSGR 1106
QY 526 SVYAYEHL 533
Db 1107 NIPAKFHV 1114

RESULT 2
US-09-491-356C-9
; Sequence 9, Application US/09491356C
; Patent No. 6566061
; GENERAL INFORMATION:
; APPLICANT: Philibert, Robert A.
; APPLICANT: Ginns, Edward I.
; TITLE OF INVENTION: IDENTIFICATION OF POLYMORPHISMS IN THE PCTG4 REGION OF XQ13
; FILE REFERENCE: 9465.6US11
; CURRENT FILING DATE: 2000-01-26
; PRIOR APPLICATION NUMBER: PCT/US99/09365
; PRIOR FILING DATE: 1999-04-29
; PRIOR APPLICATION NUMBER: 60/083,465
; PRIOR FILING DATE: 1998-04-29
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 2074
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-491-356C-9

Query Match 4.4%; Score 137.5; DB 4; Length 2074;
Best Local Similarity 18.0%; Pred. No. 0.001;
Matches 137; Conservative 118; Mismatches 226; Indels 281; Gaps 36;

QY 17 NPLQRHQSEQRRLPKIVETESTMDITIGSKQPO-----FLKSIDELAAFSVAVETF 72
Db 1215 DPLVSSAQAR-----LMQLICYPHRLDNDGPNQORIKRILKLDOW-----TW 1262
QY 73 KROFDLQKHIESIENAIKSLKESNGVLAARNNFHQPMLSPRRNVSVETTVTSOPS 132
Db 1263 ROSSLELQMIKQNTNT-----ENSLI-----ENIA-KATIEVFOQS 1299
QY 133 QIVPET-----SNKEPGRMCELMCSKGLRKYIVANTSDQAKLMEEIPSAKLAKPAK 188
Db 1300 AETGSSSGSTAGSNMPSSSKTKPVLSS--LER---SGVWLVAFLAKLFTSVQ-----GH 1348
QY 189 VLDCGKGFYQGR---AFTKESPMSSARQVSLILLESFL-----LMPDRGKGKVKIESW 240
Db 1349 VLKAGEELEKQHLGSSGRKRDQKQKMSLLSQPFSLVLTCLKQDQREGLLAS 1408
QY 241 IKDEAETAFAVAVR-----KSLMTE-----GGL-----AAAEKMDARGLL 275
```

```
Db 1409 LHSQVHOIWNRENQYLDCKPKQLMHEALKRLNLVGMFDTVQORSTQOTTEWAQLLL 1468
QY 276 LVACFGV-----PSNFRSTDLIDLIRMSGSNEIAGALKRSQFLVPMVSGIVESSIKRGMH 331
Db 1469 EIIISGVDMQSNELFTTVDMLSVLNGTLAADY-----SSISQG---1510
QY 332 BALEMVTYTFGMEKFSAAVLTSFL-----KMSKESFERAKRKAQSPKAPKEAAT-----381
Db 1511 -----SMEENKRAYMNLVKLQKDLGERQSDSLEKVHQLLPLPKQNRDVITCEPQG 1561
QY 382 -----KQLAVLSSV-----MQMETHKLDPAKELPGWQKEQIVSLEK-----DTL 422
Db 1562 SLITKGNKLAGFDSIPFKEGQLQVSTQKISF-----WELFGLKPSAPLSWAWFGTV 1615
QY 423 QLDKEM---BEKARSLMEBAALAKEMYNQOIK-RPR---LSPMEMPPVTSSSYPIY- 474
Db 1616 RVDRRVARGEQQRL-----LYHTHLRPRRAYVLEPLPLPPEDEPPAPALL 1664
QY 475 -----RDRSFPQQRDDQDEISALVSSYLGPTST 502
Db 1665 EPEKAPPEPKTDKPGAAPPSTBERKKKTKGKRSQPATKNEDYG-----MGPR 1715
QY 503 SFPHRSRRSPHYM-----VPLPHGQ-----522
Db 1716 SGYGVTVPPDLLHHPANSGISHLSYQSSMGLYTONQPLPAGPRVDYPRVRLPMQKL 1775
QY 523 -----LGRSVAYEHLAPNSYSPGHGRHLRQYSPSLVHQG---RHPLQYSPPIHQ 571
Db 1776 PTRPTYPGVLPFTMTSTMVGLPSSYKTS-----VYRQOQPTVPQGRQLRQOQSQMLQ 1831
QY 572 ---QQLP---YGIQRYVRHSPSEERY-----LGLSNQSRSP 600
Db 1832 SSVHQMTPSSSYGLQTSQSLSPSLQGYTSVSHVGLQQHTGP 1873

RESULT 3
US-09-491-356C-8
; Sequence 8, Application US/09491356C
; Patent No. 6566061
; GENERAL INFORMATION:
; APPLICANT: Philibert, Robert A.
; APPLICANT: Ginns, Edward I.
; TITLE OF INVENTION: IDENTIFICATION OF POLYMORPHISMS IN THE PCTG4 REGION OF XQ13
; FILE REFERENCE: 9465.6US11
; CURRENT APPLICATION NUMBER: US/09/491,356C
; CURRENT FILING DATE: 2000-01-26
; PRIOR APPLICATION NUMBER: PCT/US99/09365
; PRIOR FILING DATE: 1999-04-29
; PRIOR APPLICATION NUMBER: 60/083,465
; PRIOR FILING DATE: 1998-04-29
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 2023
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-491-356C-8

Query Match 4.0%; Score 123.5; DB 4; Length 2023;
Best Local Similarity 18.0%; Pred. No. 0.025;
Matches 136; Conservative 120; Mismatches 225; Indels 273; Gaps 36;

QY 17 NPLQRHQSEQRRLPKIVETESTMDITIGSKQPOFLKSIDELAAFSVAVETPKRF 76
Db 1156 DPLVSSAQAR-----LMQLICYPHRLDNDGPNQORIKRILQ-----NLDQWTMRQS 1207
QY 77 DDLCKHI-ESIENAIKSLKESNGVLAARNNFHQPMLSPRRNVSVETTVTSQSQBI 135
Db 1208 LELQMIKQTPNEMNSLE-----NIA-KATIEVFOQSAT 1243
QY 136 VPET-----SNKEPGRMCELMCSKGLRKYIYANISQAKLMEEIPSAKLAKPAK 191
```

Db 1244 GSSSGTASNPSSKTKFVLSS--LER--SGVNLVAPLIAKLPTSVQ-----GHVLK 1292  
Qy 192 CIGKPYLOGR--ATTKSPMSAPQVSLILLEGFLM-----PDRGKGVK--- 236  
Db 1293 AAGEELEKGOHLGSSRKDRQKQKMSLLSQPFSLVLTCLGQDEQREGILTSYS 1382  
Qy 237 -----IESWIKDR-----AETAAVAMKRLMTGGGL-----AAAEKMDARGILLVA 278  
Db 1353 QVHQIVNNRFDQYLDCKPKQLMHEALKRLNLVGMFDTVORSTQQTTEWAMILLEII 1412  
Qy 279 CFV-----PSNFSTDLDIRMSGNEBAGALKSQFVLPVMSGIVSSIKGNGHIAL 334  
Db 1413 ISCTVMSQNNELFTVLDMLSVLNGTLAADM-----SSISQ----- 1451  
Qy 335 EMVVTGMEKDFGAALVLTSLFKMSKE-----SPERAKRKAQSPFLAKKAAT----- 381  
Db 1452 -----SMEENKEAYNNLAK--KLQELGERQSDSLEKVRQLPLPKQTRDVTICEPOGS 1503  
Qy 382 -----KQIAVLSSV-----MQMETHKLDPAKELPGWQIKQIYSLEK-----DTLQL 424  
Db 1504 LIDTGNKNIAGPDSIFKKEGLVSTTKQKISP-----WDLFGKLKPSAPLSWGNFGTVRV 1557  
Qy 425 DKEM-----EEKARSLMEEAALAKRMYNQIK-RPR-----LSPMEMPPVTSSSYSPY--- 474  
Db 1558 DRRVARGEEOQLL-----LYHTHLRPRPRAYYLEPLPPEDEEPATLLEP 1606  
Qy 475 -----RRSFPQRDDQDEISALVSSYLGPSTSF 504  
Db 1607 EKAPEPKTKFGAAPPSTEERKKXKTKGKRSQATKTEDYG-----MGFGRS 1657  
Qy 505 PHRSRRSPXYM-----VPLPHGG----- 522  
Db 1658 PYGVTPDHLHPNPGSITHLYRQSGSLGTYNQPLPAGPRVDPRPVPLPMQKLP 1717  
Qy 523 -----LGRSVAYEHLAPNSYSPGHRLHRQYSPSLVHQO--RHPLOYSPPHGQ-- 571  
Db 1718 RPTVGVLPITMTGMLSPSSYKTS-----VYRQQPAPVQGGRLRQQLQSQGMLGQSS 1773  
Qy 572 --QQLP---YGIQVYRHSSEERYGLSNQSRP 600  
Db 1774 VHQMTSSSYGLQTSQGYTPYVS-HVGLQOHTGP 1806

RESULT 4  
US-09-308-375-2  
; Sequence 2, Application US/09308375  
; Patent No. 6300117  
; GENERAL INFORMATION:  
; APPLICANT: Genencor International, Inc.  
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms  
; FILE REFERENCE: GC394-PCT  
; CURRENT APPLICATION NUMBER: US/09/308,375  
; CURRENT FILING DATE: 1999-05-14  
; EARLIER APPLICATION NUMBER: EP9719636.4  
; EARLIER FILING DATE: 1997-09-15  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 2285  
; TYPE: PRT  
; ORGANISM: Bacillus subtilis  
US-09-308-375-2

Query Match 3.9%; Score 122.5; DB 4; Length 2285;  
Best Local Similarity 19.3%; Pred. No. 0.038;  
Matches 83; Conservative 70; Mismatches 151; Indels 125; Gaps 15;  
Qy 159 KYIVANISD---QAKLMEIPSALKAKPAKFLVDCIGKFKYQGRRAFTKSPMSAQ 215  
Db 526 KTIIFARIGNQSSIKALEIGISVKTAGGASASDLIS-----VAGKNDTLSDAQK 578  
Qy 216 VSLILLESFLMPDRGKGVKIESWIKDEABTAA-----VAWRKRL----- 256

Db 579 QNTSIGVAGIYQLSRFNAMNMFNFIQAQNAAKTAANSTGSAWSEQQKYADSLQARVNKLQN 639  
Qy 257 -MTEGGLAAERKMDARGLLLVACFGVPSNFRSTDLDIRMSG--NEIAGALKRSQFL 313  
Db 639 NTEFAIASDA-----FISDGLIEFTQAAGSLLNASTGVIKSVGFL 680  
Qy 314 VPMVSGI-----VSSIKRGMHIEALEMYVTFGMEKDFSAALV----- 351  
Db 681 PFLAAVSTATILLSKNTRTLLASSILGLTRAMGOFTLATAGLEAGMTRAAVASRVLKTAL 740  
Qy 352 -----LTSF--LKMWSKESFERAKR-KAQSPLAPKAEATKQLA 385  
Db 741 RGLLVSTLVGGFAALGALGWALESLSISFAEAKKAKDQFESQOQTNVEAITTKDSTDKLIQ 800  
Qy 386 VLSSVMQCMETHKLDPAKELPGWQIKQIV-----SLEKOTLQDKMEE 430  
Db 801 QYKELQKVKESRLTSDEQEYLVQTOQLAOTFFALVKGYDSQGNAILKTNKELEKAIEN 860  
Qy 431 KARSLSL--MEEAALAKRMYNQIKRPLSPMEMPPVTSSSYSPI--YRDRSFFSQ---R 483  
Db 861 TKEYLALKQETRDSAKTIFEDASKEIKSKDEL-----KQYKQIADYNDKGRPKWDLIA 915  
Qy 484 DDQDEISA 492  
Db 916 DDDDYKVA 924

RESULT 5  
US-09-060-410-2  
; Sequence 2, Application US/09060410  
; Patent No. 6165461  
; GENERAL INFORMATION:  
; APPLICANT: Cobb, Melanie  
; APPLICANT: Hutchinson, Michele  
; APPLICANT: Chen, Zhu  
; APPLICANT: Bertman, Kevin  
; TITLE OF INVENTION: TAO PROTEIN KINASES AND METHODS OF USE  
; TITLE OF INVENTION: THEREFOR  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/060,410  
; FILING DATE: 14-APR-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 860098.421  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1001 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-060-410-2

Query Match 3.9%; Score 120.5; DB 3; Length 1001;  
Best Local Similarity 19.8%; Pred. No. 0.015;







; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.  
 ; STREET: P.O. Box 828  
 ; CITY: Bloomfield Hills  
 ; STATE: MI  
 ; COUNTRY: USA  
 ; ZIP: 48303  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/742,923A  
 ; FILING DATE: No. 5969611ember 1, 1996  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Smith, DeAnn F.  
 ; REGISTRATION NUMBER: 36683  
 ; REFERENCE/DOCKET NUMBER: 2115-00869DVC  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (810) 641-1600  
 ; TELEFAX: (810) 641-0270  
 ; INFORMATION FOR SEQ ID NO: 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 816 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-742-923A-6

Query Match 3.7%; Score 113.5; DB 2; Length 816;  
 Best Local Similarity 21.1%; Pred. No. 0.052;  
 Matches 100; Conservative 76; Mismatches 190; Indels 109; Gaps 20;

Qy	24	QSEORRELEPKIVETES-----TSMDTIGOSKQPOFLKSIDELAAPSVAVERFK 73
Db	218	QLEERNSLOQDDEMEAKQNLERHISTLNQLSDSK-----KLODFASVLEAEKGK 272
Qy	74	RQPDLOKHIESINAIDSKLENGVVLAAARNNNHQPMLSPRRNVSVETTVVSQPSQ 133
Db	273	KRF---QKEIENLTQOYEEKAAAYDKLETKN-----RLOQELDDLVDLDNQR 318
Qy	134	EIVPETSNGPEGRMCELMCSKGLRKVIYANISD--QAKLMBEIPSAKLAK--BPAKFV 189
Db	319	QLVSNLEKKQR--KFDOLLAEBKNISSKYADERDRAEAAREKETKALSARALEEA--- 373
Qy	190	LDICIGFYLOGRRAFTKESPMSSARQVSLILLESFLMLPDRGKGKVKIESWIKDEATAA 249
Db	374	-----LEAKEBELRTNKLKAEEDLVSSKD-----DVGKNVHELEK-SKRALETQM 419
Qy	250	VAMRKRL-MTEGGLAAAEKMDARGLLLVACFGVPSNFRSTDLLDIRMSGNEITAGALK 308
Db	420	BEKMTQLELEDELOASE--DAK-LRLEVNMQALKQOGER-----DLQARDEQNE-----EK 468
Qy	309	RSQFLVPMVSVGIVESSIKRGMHIEALEMVTYFGMEDKFSAAVLVTSFLKMSKESPERAKR 368
Db	469	RRQ-----LQRLH-----EYETELEDERNERALEAAAKKKLEGDLKDLLEL 509
Qy	369	KAQSPLAFKEAATKQLAVLSSVMQ-----CMETHKLDPAKELPGMQ 409
Db	510	QADSAIKGREEAIKQLRKLOAQKDFORELEDARASRDEIFATAKENEKAKSLEADLMQ 569
Qy	410	IKEQIVSLEKDTLQDKEMEKEKARSLSLMBEAAALAKRMVNNQIKR---PRLSPME 461
Db	570	LQEDLAAAEARAKQADLEKELEAEELA-----SSLSGRNALQDEKRRLEARIAQLE 620

RESULT 11  
 US-08-533-306A-4  
 ; Sequence 4, Application US/08533306A  
 ; Patent No. 5837457  
 ; GENERAL INFORMATION:

; APPLICANT: Liu, Pu  
 ; APPLICANT: Collins, Francis S.  
 ; APPLICANT: Siciliano, Michael J.  
 ; APPLICANT: Claxton, David  
 ; TITLE OF INVENTION: Markers for Detection of Chromosome 16  
 ; TITLE OF INVENTION: Rearrangements  
 ; NUMBER OF SEQUENCES: 14  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.  
 ; STREET: P.O. Box 828  
 ; CITY: Bloomfield Hills  
 ; STATE: MI  
 ; COUNTRY: USA  
 ; ZIP: 48303  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/533,306A  
 ; FILING DATE: September 25, 1995  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Smith, DeAnn F.  
 ; REGISTRATION NUMBER: 36683  
 ; REFERENCE/DOCKET NUMBER: 2115-00869COB  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (810) 641-1600  
 ; TELEFAX: (810) 641-0270  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 885 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-533-306A-4

Query Match 3.7%; Score 113.5; DB 2; Length 885;  
 Best Local Similarity 21.1%; Pred. No. 0.06;  
 Matches 100; Conservative 76; Mismatches 190; Indels 109; Gaps 20;

Qy	24	QSEORRELEPKIVETES-----TSMDTIGOSKQPOFLKSIDELAAPSVAVERFK 73
Db	287	QLEERNSLOQDDEMEAKQNLERHISTLNQLSDSK-----KLODFASVLEAEKGK 341
Qy	74	RQPDLOKHIESINAIDSKLENGVVLAAARNNNHQPMLSPRRNVSVETTVVSQPSQ 133
Db	342	KRF---QKEIENLTQOYEEKAAAYDKLETKN-----RLOQELDDLVDLDNQR 387
Qy	134	EIVPETSNGPEGRMCELMCSKGLRKVIYANISD--QAKLMBEIPSAKLAK--BPAKFV 189
Db	388	QLVSNLEKKQR--KFDOLLAEBKNISSKYADERDRAEAAREKETKALSARALEEA--- 442
Qy	190	LDICIGFYLOGRRAFTKESPMSSARQVSLILLESFLMLPDRGKGKVKIESWIKDEATAA 249
Db	443	-----LEAKEBELRTNKLKAEEDLVSSKD-----DVGKNVHELEK-SKRALETQM 488
Qy	250	VAMRKRL-MTEGGLAAAEKMDARGLLLVACFGVPSNFRSTDLLDIRMSGNEITAGALK 308
Db	489	BEKMTQLELEDELOASE--DAK-LRLEVNMQALKQOGER-----DLQARDEQNE-----EK 537
Qy	309	RSQFLVPMVSVGIVESSIKRGMHIEALEMVTYFGMEDKFSAAVLVTSFLKMSKESPERAKR 368
Db	538	RRQ-----LQRLH-----EYETELEDERNERALEAAAKKKLEGDLKDLLEL 578
Qy	369	KAQSPLAFKEAATKQLAVLSSVMQ-----CMETHKLDPAKELPGMQ 409
Db	579	QADSAIKGREEAIKQLRKLOAQKDFORELEDARASRDEIFATAKENEKAKSLEADLMQ 638
Qy	410	IKEQIVSLEKDTLQDKEMEKEKARSLSLMBEAAALAKRMVNNQIKR---PRLSPME 461
Db	639	LQEDLAAAEARAKQADLEKELEAEELA-----SSLSGRNALQDEKRRLEARIAQLE 689

```

QY 369 KAQSFLAKEAATQTLAVLSVWQ-----CMETHKLDPAKELFGW 409
DB 579 QADSAIKREIEAIFQIKRLQAKQMDQFQRELDARASRDEIFATAKENKKAKSLEADLMQ 638
QY 410 IKEQIVSLKEDTLDKREMEKEKARSLSLMEEAALAKRMVYNQOIKS---PRLSPME 461
DB 639 LQEDLAAARARKQADLEKEELABELA---SSLSGRNALQDEKARLEARIAQLE 689

RESULT 13
US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 559919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; ATTORNEY/AGENT INFORMATION:
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO. 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
; US-08-353-700-1

Query Match 3.6%; Score 112; DB 1; Length 3248;
Best Local Similarity 19.5%; Pred. No. 0.76;
Matches 115; Conservative 105; Mismatches 219; Indels 150; Gaps 24;

QY 11 QTTTANPLIQHSGEORRR-----ELPKIVETESTSMDITTGSKQKQFQFKSDE- 61
DB 2530 EKSSTAMEMLQTQLKELNERVAALHNDQACKAKE-QNLSQVECLEKLAQLLQGLDEA 2588
QY 62 -----LAAFSVAVEFKQFDDQLKHIESIENAI-----SKL----- 94
DB 2589 KNNYIVLQSSVKGLIQVEDGKQKLEKDEEISRLKXQIQDQQLVSKLSQVEGEHQLWK 2648
QY 95 ESVGTVLAARN--NNFHQPMLSPPRNNSVETTVTSQDSQIVETSNKPEGGRMCELM 152
DB 2649 EQN-----LELRNLTVLEEQKTVQTSKNASLQDTLEVLOSSYK---NLNELELTNMDKMS 2702
QY 153 CSKGLRKIYIVANSQAKLMEEIPSAKLAKPEAKFVLDICG-KFYLGRRAFTKESPMS 211

```

Db 2703 FVEKVNKM---TAKETELQREMH---EMAQKTAELOEELSCEKNRLAGELQLLEIEIKS 2755  
 Qy 212 SARQVSLILLESFLL---MPDRGKGVKIESWIKDEA----- 245  
 Db 2756 SKDQLKELTLENSLKSLDCMHKQVKEGKVRBEIAEYQLRLHEAEKHKQALLDITNK 2815  
 Qy 246 --ETAAVAWKRLMTGGLAAAKMDARGILLLVACFGVPSFRSTDLLDIRSGSNEI 303  
 Db 2816 QYVEIQTVREKLTSEKCLSSQKLE-----IDLLK-SKKEEL 2852  
 Qy 304 AGALKRSQFLVPMVSGIVESSIKRGHIEBALENVYTFGMEKFSAA---LVLTSFLKMSK 360  
 Db 2853 NNSLKAT-----TQILEELKXTKM--DNLYNQVKKENERAQGMKLLIKSCQKLE 2903  
 Qy 361 ESFERAKRAQSLAPKEAATKQLAVLSSVMQCMETHKLDPAK--ELPGWQIKQIVSLE 418  
 Db 2904 E-----KEILOKELSQLOAAQEKQKTGTVMKTVDLT-TEIKELKETLE 2947  
 Qy 419 KDTLOLDKEMEERKARSLMEEAALAKMYNQOI-----KRPRLSPMEMPVTVSSS 469  
 Db 2948 EKTKEADYLDKYCSLLISHEKLEKAKEMLETOVAHLCSQSQKDSRGSPLLGPVVPGPS 3007  
 Qy 470 YSPIYDRSRFPS---QRDDQDDEISALVSSYLGPS-----TSPHRSRRS 511  
 Db 3008 PIPSVTEKRLSSGQNKASGRQRSSGIWENGRGPTPATPESFSKSKKA 3056

RESULT 14  
 PCT-US95-16216-1  
 ; Sequence 1, Application PC/TUS9516216  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Yen, Timothy J.  
 ; APPLICANT: Rattner, Jerome B.  
 ; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently  
 ; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Pann, Dorfman, Herrell and Skillman  
 ; STREET: 1601 Market Street Suite 720  
 ; CITY: Philadelphia  
 ; STATE: PA  
 ; COUNTRY: USA  
 ; ZIP: 19103-2307  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US95/16216  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/353,700  
 ; FILING DATE: 09-DEC-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Reed, Janet E.  
 ; REGISTRATION NUMBER: 36,252  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (215) 563-4100  
 ; TELEFAX: (215) 563-4044  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3248 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: not relevant  
 ; TOPOLOGY: not relevant  
 ; MOLECULE TYPE: protein  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 PCT-US95-16216-1

Query Match 3.6%; Score 112; DB 5; Length 3248;

Best Local Similarity 19.5%; Pred. No. 0.76;  
 Matches 115; Conservative 105; Mismatches 219; Indels 150; Gaps 24;  
 Qy 11 QPTTANPLQLHQSHQSEORRR-----ELPKIVETESTMDITIGOSKQPFLLKSID- 61  
 Db 2530 EKSSSTAMEMLTQKLNELNVAALHNDQEAACKAKE-QNLSQVECLELEKAQLQLGLDEA 2588  
 Qy 62 -----LAAPSVAVETFKRQDDDLQKHIESIENAIID-----SKL----- 94  
 Db 2589 KNNYIVLOSVMKGLQOEVDGKQLEKDEEISRLKQIQOEQLVSKLSQVEGHEQLNK 2648  
 Qy 95 EBSGVVLAARN--NNFHOPMLSPPPNNVSVTTVTVSQSQEIVPETSNNKPEGGMCELM 152  
 Db 2649 EQN---LELRNLTVELEQKIQVLOSKNASLQDTLLEVLSQSYK--NLNLELELTCKMDKMS 2702  
 Qy 153 CSKGLKYYIYANISQAKLMEHIEPSALKLAKPAKFVLDICIG-KFYLCGRRAFTKESFMS 211  
 Db 2703 FVEKVNKM---TAKETELQREMH---EMAQKTAELOEELSCEKNRLAGELQLLEIEIKS 2755  
 Qy 212 SARQVSLILLESFLL---MPDRGKGVKIESWIKDEA----- 245  
 Db 2756 SKDQLKELTLENSLKSLDCMHKQVKEGKVRBEIAEYQLRLHEAEKHKQALLDITNK 2815  
 Qy 246 --ETAAVAWKRLMTGGLAAAKMDARGILLLVACFGVPSFRSTDLLDIRSGSNEI 303  
 Db 2816 QYVEIQTVREKLTSEKCLSSQKLE-----IDLLK-SKKEEL 2852  
 Qy 304 AGALKRSQFLVPMVSGIVESSIKRGHIEBALENVYTFGMEKFSAA---LVLTSFLKMSK 360  
 Db 2853 NNSLKAT-----TQILEELKXTKM--DNLYNQVKKENERAQGMKLLIKSCQKLE 2903  
 Qy 361 ESFERAKRAQSLAPKEAATKQLAVLSSVMQCMETHKLDPAK--ELPGWQIKQIVSLE 418  
 Db 2904 E-----KEILOKELSQLOAAQEKQKTGTVMKTVDLT-TEIKELKETLE 2947  
 Qy 419 KDTLOLDKEMEERKARSLMEEAALAKMYNQOI-----KRPRLSPMEMPVTVSSS 469  
 Db 2948 EKTKEADYLDKYCSLLISHEKLEKAKEMLETOVAHLCSQSQKDSRGSPLLGPVVPGPS 3007  
 Qy 470 YSPIYDRSRFPS---QRDDQDDEISALVSSYLGPS-----TSPHRSRRS 511  
 Db 3008 PIPSVTEKRLSSGQNKASGRQRSSGIWENGRGPTPATPESFSKSKKA 3056

RESULT 15  
 US-08-328-254-6  
 ; Sequence 6, Application US/08328254  
 ; Patent No. 5710022  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Zhu, Xueliang  
 ; APPLICANT: Lee, Wen-Hwa  
 ; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein  
 ; NUMBER OF SEQUENCES: 8  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Campbell and Flores  
 ; STREET: 4370 La Jolla Village Drive, Suite 700  
 ; CITY: San Diego  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 92122  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/328,254  
 ; FILING DATE: 24-OCT-1994  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/141,239  
 ; FILING DATE: 22-OCT-1993  
 ; ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-CJ 1191  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2482 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-328-254-6

Query Match 3.5%; Score 110; DB 1; Length 2482;  
Best Local Similarity 18.9%; Pred. No. 0.76;  
Matches 140; Conservative 120; Mismatches 272; Indels 208; Gaps 30;  
QY 11 QPTTANPLLQHQHQRR-----ELPKIVETSTMDITIGSQKQFQPKSIDE-61  
DB 1802 EKSSTAMEMLOTQKELNERNVAALHNDQEAACKAE-QNLSSQVECLEKAQLQLGLDEA 1860  
QY 62 -----LAAPSVAVETFKRQFDDLQKHIESIENAIID-----SKL-----94  
DB 1861 KNYIVLQSSVNGLTQEVEDGQKLEKDEHSIKKNQIQDQEQLVSKLSQVEGEHQLMK 1920  
QY 95 ESNQVVLAAARN--NNFHQPLSPPRNNVSVETTVTSQSQEIIVFETSNKPEGRMCCLM 152  
DB 1921 EQN---LELRNLTVLEQKIQVLOSKNASLQDTLEVLQSSYK---NLENELELTQMDKMS 1974  
QY 153 CSKGLRKIYANISDQAKLMEIIFSAKLAKPAKFVLDICG-KFYLGRRRAFTKESPMs 211  
DB 1975 FVEKNKMM---TAKETELQREMH---EMAQTAELQELSGEKNRLAGELQLLLEIKS 2027  
QY 212 SARQVSLILIESFLL---MPDRGKGKVKIESWIKDEA-----245  
DB 2028 SKDQLKELTLENSLKSLDCMHWQDQVEKGEKVREIETAYQLRLHEAEKKHQALLDITNK 2087  
QY 246 --ETAAVAKRELMTTEGLAAAKXMDARGLLLLVACFGVPSNPFSTDLDDLIRMSGSEI 303  
DB 2088 QYEVETIQTREKLTSEKLSQKLE-----IDLK-SKEEL 2124  
QY 304 AGALKRSQFLVPMVSGIVESIKRGHIEALEMVTYFGMEDKFSAA---LVLTSTFLAMSK 360  
DB 2125 NNSLKAT-----TQILEELKTKV--DNLKYVNLKKNENERAQGKMKLLIKCKQLEE 2175  
QY 361 ESFERAKRAQSPAPAKAAIKQALVLSVMQCMETHKLDPAK--ELPGHQIKEQIVSLR 418  
DB 2176 E-----KEILOKLSQQAQKOKGTGVTMDTKYDELT-TEIKELKETLE 2219  
QY 419 KDTQLDKEMEERKARSLSLMEEAALAKEMYNQOI-----KRPRLSPMEMPPVTSSS 469  
DB 2220 EKTKEADEVLKVCYSLISHEKLEKAKEMLTQVAHLCSQQSKQDSRGSPLLGVPVFGPS 2279  
QY 470 YSPIVRDRSFPs---QRDDQDEISALVSSYLGPS-----TSFPHRSRRS-----511  
DB 2280 PIPSVTEKRLSSGQNKASGRSSGIWENGSGGTPATPESFSKSKKAVMSGIHPAEDT 2339  
QY 512 -----PEVM---VPLPHGG-----LGRSVYA-----YEH 532  
DB 2340 EGTEFEPEGLPEVVKGFADITGKTSTYILRTTMTATSPRIAAQKALSPSLGKEN 2399  
QY 533 LAPNSYSPGHRLHRQVSPSLVHQRHPLQVSPPIHGQQQLPYGIORVYVHSPSEERYL 592  
DB 2400 LAESSKPTAGGSR-----SQVKVAQRSPVDSGTTILREPTTKSVFVNNLPERSPTDPRE 2454  
QY 593 GLSNQRS---PRNSSLDEK 609  
DB 2455 GURVRRGLVPSPKAGLESK 2474

Search completed: August 14, 2003, 10:41:27  
Job time : 33 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

QM protein - protein search, using sw model

Run on: August 14, 2003, 10:40:05 ; Search time 59 seconds  
(without alignments)  
1352.220 Million cell updates/sec

Title: US-09-890-475-1

Perfect score: 3104

Sequence: 1 MSNYPPTVAQPTTANPLL.....RYLGLSNORSPRSNSLDPX 609

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 492763 seqs, 131003257 residues

Total number of hits satisfying chosen parameters: 492763

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	144	4.6	725	11	US-09-978-309A-47
2	132.5	4.3	888	11	US-09-893-519A-73
3	129	4.2	771	15	US-10-204-887-110
4	127.5	4.1	786	11	US-09-291-417-23
5	127.5	4.1	1001	11	US-09-291-417-31
6	127.5	4.1	1005	15	US-10-177-293-254
7	127.5	4.1	3256	10	US-09-919-172-98
8	127.5	4.1	3256	11	US-09-919-039-21
9	124.5	4.0	1203	15	US-10-097-340-43
10	124	4.0	426	15	US-10-156-761-12795
11	122.5	3.9	2285	10	US-09-332-183A-2
12	115.5	3.7	1001	15	US-10-128-714-3240
13	114.5	3.7	862	15	US-10-177-293-391
14	114.5	3.7	1938	15	US-10-171-311-164
15	114.5	3.7	1945	11	US-09-927-597-2

16 114.5 3.7 1972 15 US-10-171-311-162 Sequence 162, App  
17 114.5 3.7 1979 11 US-09-927-597-4 Sequence 4, Appl  
18 113 3.6 647 9 US-09-841-132-436 Sequence 436, App  
19 112.5 3.6 892 12 US-10-205-219-102 Sequence 102, App  
20 112.5 3.6 914 15 US-10-128-714-8595 Sequence 8595, App  
21 111.5 3.6 1069 15 US-10-146-473-77 Sequence 77, Appl  
22 111.5 3.6 26926 10 US-09-759-508B-2 Sequence 2, Appl  
23 111.5 3.6 2139 9 US-09-727-384-6 Sequence 6, Appl  
24 110.5 3.6 2139 15 US-10-023-219-4 Sequence 4, Appl  
25 110 3.5 1101 15 US-10-177-293-106 Sequence 106, App  
26 109.5 3.5 660 14 US-10-007-693-139 Sequence 139, App  
27 109 3.5 578 9 US-09-159-469-50 Sequence 50, Appl  
28 109 3.5 578 9 US-09-798-042-50 Sequence 50, Appl  
29 108.5 3.5 748 9 US-09-864-761-43244 Sequence 43244, A  
30 108 3.5 664 9 US-09-823-038A-48 Sequence 48, Appl  
31 108 3.5 1867 11 US-09-824-574-5 Sequence 5, Appl  
32 107.5 3.5 749 10 US-09-833-205-4 Sequence 4, Appl  
33 107.5 3.5 1233 11 US-09-291-417-89 Sequence 89, Appl  
34 106.5 3.4 489 9 US-09-876-889-350 Sequence 350, App  
35 106 3.4 1618 10 US-09-963-875-1 Sequence 1, Appl  
36 106 3.4 1618 15 US-10-136-891-2 Sequence 2, Appl  
37 106 3.4 1618 15 US-10-120-887-1 Sequence 1, Appl  
38 105.5 3.4 945 9 US-09-745-763-191 Sequence 191, App  
39 105 3.4 594 12 US-10-137-870-10 Sequence 10, Appl  
40 105 3.4 594 12 US-10-140-018-10 Sequence 10, Appl  
41 105 3.4 594 12 US-10-140-021-10 Sequence 10, Appl  
42 105 3.4 594 12 US-10-140-274-10 Sequence 10, Appl  
43 105 3.4 594 12 US-10-140-471-10 Sequence 10, Appl  
44 105 3.4 594 12 US-10-140-807-10 Sequence 10, Appl  
45 105 3.4 594 12 US-10-140-922-10 Sequence 10, Appl

#### ALIGNMENTS

#### RESULT 1

US-09-978-309A-47

; Sequence 47, Application US/09978309A

; Publication No. US20030100490A1

; GENERAL INFORMATION:

; APPLICANT: Cruz, Tony

; APPLICANT: Pastrek, Aleksandra

; APPLICANT: Turley, Eva A.

; TITLE OF INVENTION: Compositions and Methods for Treating Cellular Response to Injury and Other Proliferating Cell Disorders Regulated by Hyaladherin and Hyaluronans

; TITLE OF INVENTION: Hyaladherin and Hyaluronans

; FILE REFERENCE: 033352-010

; CURRENT APPLICATION NUMBER: US/09/978,309A

; CURRENT FILING DATE: 2002-04-04

; PRIOR APPLICATION NUMBER: US 09/685,010

; PRIOR FILING DATE: 2000-10-05

; PRIOR APPLICATION NUMBER: US 09/541,522

; PRIOR FILING DATE: 2000-04-03

; PRIOR APPLICATION NUMBER: US 60/127,457

; PRIOR FILING DATE: 1999-04-01

; NUMBER OF SEQ ID NOS: 84

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 47

; LENGTH: 725

; TYPE: PRT

; ORGANISM: Homo sapien

US-09-978-309A-47

Query Match 4.6%; Score 144; DB 11; Length 725;

Best Local Similarity 19.6%; Pred. No. 0.00094;

Matches 109; Conservative 91; Mismatches 226; Indels 130; Gaps 20;

QY 34 KIVETESTMDITIGSKQPOFLKSIDELAASFVAVETFKQFDLDKQHSIENADS- 92

DB 216 KIVSIEKEKID---EKSETEKLEVEIEBISCSAQVEKYLKIDIAQLEENLKEKNDEISL 272

QY 93 --KLESGNVLA-----ARNNNFQPMLSPPRRNVSVETTTVVSQ 130

Db 273 KOSLEDNIVLSQVEDLNVKQCLLEKEDHVNRRNHNENNAEMQN--LEOKFILEQ 330  
Qy 131 PSOEIVPETSNEPGRMCELMCSKGLRKYIYANISDQAKLMEIIPSAKLAKPAKFAVL 190  
Db 331 REHEKLOKELQIDSLLOQEKELSSSLHQL-----CSQCEWVKEKNLFEELKQTL 383  
Qy 191 DCIGKPYL---QORRAFTKESPMSSARQVSLILLESFLLMPDRGKGVKIESWIKDEAET 247  
Db 384 DELDKLQKEQEAERLVKQLEBEAKSRAEELKLEKL-----KGK-----EABLEX 430  
Qy 248 AAVAKRLMTTEGGLAAAEKMDARGILLILVACFVPSNFRS-----TDL 291  
Db 431 SNA-----HTCATLLQEKYDS-----MVSLEDVTAQFESYKALTASEIEDKLESSL 481  
Qy 292 LDIRMGSN-----EIGALKRSQFLVPMVSGIVESSIKRGHIEALEMVTYFGMEDK 345  
Db 482 QEKAAKAGNAEDVQOILATESNQEVYRMLDLQTKSALKETEIKEITVSFLQKITD- 540  
Qy 346 FSAALVLTSLFKMSKESPERAKKQSPAFKBAATKOLAVLSSVMQCMETHKLDPAKEL 405  
Db 541 -----LQNLQKQEDFRKQLEDEGRKAEKENTAEI-----TEEI 577  
Qy 406 PGWQ-KEQIVSLEKD-TLQLDK-EMBEKARSLSMEEAALAKRMYN-----Q 450  
Db 578 NKWRLLYEELYNKTKPFQQLDAFEVEKQA-----LLNEHGAQEQINKIRDSYAKLLCHQ 633  
Qy 451 QIKRPLSPMEMPVPTSSSYSPYDRSPPSORDDDQDEISALVSSYLQ-----PSTSEF 505  
Db 634 NLKQKTKHVVLKDENSQKSEVSKURCOLANKQSETKQBELNVLGKIKHFPDPKAFH 693  
Qy 506 HRSRSPPEYMPVLPFHG 521  
Db 694 HESKENFALKTPLKEG 709

RESULT 2

US-09-893-519A-73  
; Sequence 73, Application US/09893519A  
; Publication No. US20030027243A1  
; GENERAL INFORMATION:  
; APPLICANT: ANADYS PHARMACEUTICALS, INC.  
; APPLICANT: THOMPSON, Craig  
; APPLICANT: MOORE, Jeffrey  
; APPLICANT: BUURMAN, Ed T.  
; APPLICANT: BRADLEY, John  
; APPLICANT: DESILVA, Tamara  
; APPLICANT: HARRIS, Sandra  
; APPLICANT: KOMARNITSKY, Svetlana  
; APPLICANT: MENDILLO, Marc  
; APPLICANT: MOORE, Daniel  
; APPLICANT: MCCOY, Melissa  
; APPLICANT: SANDERSON, Karen  
; APPLICANT: HAQ, Tariq  
; APPLICANT: ZHU, Shuhao  
; APPLICANT: LONG, Fan  
; APPLICANT: DAVIDOV, Eugene  
; TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE  
; FILE REFERENCE: 0342/IG348-US2  
; CURRENT APPLICATION NUMBER: US/09/893,519A  
; CURRENT FILING DATE: 2001-06-28  
; PRIOR FILING DATE: 2001-06-28  
; PRIOR FILING DATE: 2000-06-29  
; PRIOR APPLICATION NUMBER: US 60/215,164  
; PRIOR APPLICATION NUMBER: US 60/224,457  
; PRIOR FILING DATE: 2000-08-10  
; NUMBER OF SEQ ID NOS: 146  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 73  
; LENGTH: 888  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Corresponds to SEQ ID NO: 146

; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: Human Genbank/BA088763  
; DATABASE ENTRY DATE: 2000-01-06  
; RELEVANT RESIDUES: (1)...(888)  
US-09-893-519A-73  
  
Query Match 4.3%; Score 132.5; DB 11; Length 888;  
Best Local Similarity 20.0%; Pred. No. 0.014;  
Matches 108; Conservative 94; Mismatches 206; Indels 131; Gaps 23;  
  
Qy 16 ANPLQRQSQRRRELPAKIVETESTMDITIGOSQOPFLKSIDELAAAFSAVAFKQ 75  
Db 410 SNGALSTEEEREEEMKOM-EVYRSHSKFMKNKIGQVK-ELSRKDTELLALQTKLETNQ 467  
Qy 76 FPDLOKHIESIENAIKIDSLGNSGVVLAARNNNHQPMLSPNNVSVETTVTVSQSOEI 135  
Db 468 FSDSKOHIEVLEKSLUTAK-EORAAAIL-----QTEVDALRLLEKETM-UNKTKQI 517  
Qy 136 VPETENKPEGRMCELMCSKGLRKYIYANISDQAKLMEIIPSAKLAKPAKFAVLDCIG 195  
Db 518 QDMAEEK-GTQAGEIHLKMDLVKERVNLQKKTENLQEQLR----- 560  
Qy 196 FYLQORRAFTKESPMSSARQ-----VSLILLESFLLMPDRGKGVKIESWIK 242  
Db 561 -----DREKQMSLUKERVKSLQADTNTDTALTTLLEALAEKERTIERLK-BQDR 610  
Qy 243 DEATAAVAWAKRLMTTEGGLAAAEKMDARGILLILVACFVPSNFRSTDLLD-----IRMS 298  
Db 611 DERE-----KQBEIDN-----YKDLKOLKEKVSLLQGLDSEKASLLDLKEHASSLA 658  
Qy 299 GNEIAGALKRSQFLVPMVSGIVESSIKRGHIEALEMVTYFGMEDKFSALVITS----- 354  
Db 659 SDESSEKQAQAEVDRLLLELKEVENEKNDKDKIAELSLTSRQVKDQNKVKVANKHKEQV 718  
Qy 355 FLKMSKESPERAKRK-----ACSLAPAKKAATKQL-----AVLSSVMQCM-----E 395  
Db 719 EYKSAQMLEEARREDNLDSSQOLQVEELLMAWEXKQLESMAKLSSTQOSLAKE 778  
Qy 396 TH-----KLDPAKELPGWQIKQIVSLEKDTLQLDKEMEERKARSLSM-----EEA 441  
Db 779 THLTNLRAERRKHL-----BEVLEMQEALL-----AAISEKDANIALLELSSSKKKTQEEV 830  
Qy 442 ALAKEMYN---OOIKRPLSPMEMPVPTSSSYSPYDRSPPSORDD-----DQDE 489  
Db 831 AALKREKORLVQQLAQOQONRKUL-----MADNYEDDHFKSSHSHSNQTNHKSPODQE 882

RESULT 3

US-10-204-887-110  
; Sequence 110, Application US/10204887  
; Publication No. US20030124569A1  
; GENERAL INFORMATION:  
; APPLICANT: INCYTE GENOMICS, INC.  
; APPLICANT: PANZER, Scott R.  
; APPLICANT: SPIRO, Peter A.  
; APPLICANT: BANVILLE, Steven C.  
; APPLICANT: SHAH, Purvi  
; APPLICANT: CHALUP, Michael S.  
; APPLICANT: CHEN, Simon C.  
; APPLICANT: CHEN, Alice  
; APPLICANT: D'SA, Steven A.  
; APPLICANT: AMSHEY, Stefan  
; APPLICANT: DAHL, Christopher R.  
; APPLICANT: DAM, Tam C.  
; APPLICANT: DANIELS, Susan E.  
; APPLICANT: DUFOUR, Gerard E.  
; APPLICANT: FLORES, Vincent  
; APPLICANT: FONG, Willy T.  
; APPLICANT: GREENAWALT, Lila B.  
; APPLICANT: HILLMAN, Jennifer L.  
; APPLICANT: JONES, Anissa L.  
; APPLICANT: LIU, Tommy F.  
; APPLICANT: ROSEBERRY, Ann M.

APPLICANT: ROSEN, Bruce H.  
APPLICANT: RUSSO, Frank D.  
APPLICANT: STOCKREHER, Theresa K.  
APPLICANT: DAFFO, Abel  
APPLICANT: WRIGHT, Rachel J.  
APPLICANT: YAP, Pierre E.  
APPLICANT: YU, Jimmy Y.  
APPLICANT: BRADLEY, Diana L.  
APPLICANT: BRATCHER, Shawn R.  
APPLICANT: CHEN, Wensheng  
APPLICANT: COHEN, Howard J.  
APPLICANT: HODGSON, David M.  
APPLICANT: LINCOLN, Stephen E.  
TITLE OF INVENTION: SECRETORY MOLECULES  
CURRENT FILING DATE: 2002-08-21  
CURRENT FILING DATE: 2002-08-21  
PRIOR APPLICATION NUMBER: 60/185,215; 60/185,216; 60/205,232; 60/205,323; 60/205,287;  
60/205,324; 60/205,286  
PRIOR FILING DATE: 2000-02-24; 2000-02-24; 2000-05-16; 2000-05-17; 2000-05-17;  
2000-05-17; 2000-05-17  
NUMBER OF SEQ ID NOS: 159  
SOFTWARE: PERL Program  
SEQ ID NO 110  
LENGTH: 771  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID No. US20030124569A1 LI:350243.2.orf2:2000MAY01  
US-10-204-887-110

Query Match 4.2%; Score 129; DB 15; Length 771;  
Best Local Similarity 19.0%; Pred. No. 0.022;  
Matches 139; Conservative 87; Mismatches 289; Indels 216; Gaps 27;

QY 2 SNYPPTAAQTTTANPLQRHQSEQRRELPKIVET-----ESTSMDTITGCSQ 52  
DB 29 SRQPSLAASPGSSGLTAVQAAMPYSPQKIQPTAHCYGRTFHSQPCLDSSQCEKE 88  
QY 53 POFILKSI-----DELAASFVAVETFKQDFDLOKHIESIENALDSKLENGVVL 101  
DB 89 RQIEGIFKGTCKSDPDWTLVEQWHSYSPFRESSNGMKMEGLLNGSDPHQSLASI 148  
QY 102 AARNNHQPMSPRNVSVETTVTSQSQEIVP---ETSNKPEGRMCELMCSKGLR 158  
DB 149 KAEADKIYFTDNAPSIGSSKLENTTPTQPLPHVVTQNGAAS-----SVKN 201  
QY 159 KYIYANISDQAK---LMEEIPS--ALKLAKEPAKFVL-----DCIGKFLYQORRAFTKE 207  
DB 202 SPAYSDISDAGEDGEGKVDYVSKSDASQLVKEGAKTLFPFPQSKSDSPYQGFESYSP 261  
QY 208 S-----PMSARQVSLILLESFLMPDRGKGVKIESWIKDEAETAAVAWKRLM 257  
DB 262 SYAOSSPGALPSSQAG-----VESQALTKRDEEPESIEGVKND----- 302  
QY 258 TEGGLAAAEKMDARGLLILVACFGVPSNFRSTDLDIRMSGNEIAGALKRQFLVPMV 317  
DB 303 -----ICEKKPELSSSOQSPVIOQRPNMYMQSLYNYQYAVVP 342  
QY 318 SGIVESSIKRGHIEALEMVTTFQMEKFAALVLTFLKMSKESFERAKR-----KAQSP 373  
DB 343 YGYSDQSYH--THLLSTNTAYRQYEE-----QOKRQSLQEQQGVYDKKAENG 388  
QY 374 LAFKEAATKO-----LAVLSVMOCMET-----HKLDPAKE 404  
DB 389 LKREPAALKKEWKQKPSIPPTILTAPSLTDLVKSQPKAKEPGADPAKSVIIFKLDSSK 448  
QY 405 LFGMQIKQIVSLEKOTIQLDKEMEKARSLSIMEAALAKRMVNOQIKPRPSPVEMPP 464  
DB 449 LPG-QAPEGKLVK-SDAGHLSKEASEAKTGAECRQAEMDPIWYRQEAEPFMTWYVP- 506  
QY 465 VTSSYSPI-----YRDRSPSQEDDDQDDSI-----SALVS 495

DB 507 ---AKYSDIKSEDERWKEERDRKLEERSRSKDSVPKEDGKSTSSDCKLPTSESRIGS 563  
QY 496 SYLGPSTSP-----HRSRSPWVWPLPHGGLGRSVYAEHLAPNSYSPGH-GHRLH 547  
DB 564 KEPRPSVHVVSPLTQHS-----YIPYMG-----YSIS-----QSYDPNHSYR-- 605  
QY 548 RQYSPSLVHQRHPLQYSPPIHQOQOLPYGIO-----RYVRHSPSPERYLG 593  
DB 606 ---SMPAVMMQNYPGSYLPSSYSFS--PYGSKVSGEDADKARASPSVTCSSSESKALD 660  
QY 594 LSNQSRPRNS 604  
DB 661 ILQOCHASHYKS 671

RESULT 4  
US-09-291-417-23  
; Sequence 23, Application US/09291417A  
; Publication No. US20030050230A1  
; GENERAL INFORMATION:  
; APPLICANT: PLOWMAN, GREGORY  
; APPLICANT: MARTINEZ, RICARDO  
; APPLICANT: WYTE, DAVID  
; TITLE OF INVENTION: STE20-RELATED PROTEIN KINASES  
; FILE REFERENCE: 240/300  
; CURRENT APPLICATION NUMBER: US/09/291,417A  
; CURRENT FILING DATE: 1999-04-13  
; EARLIER APPLICATION NUMBER: US 60/081,784  
; EARLIER FILING DATE: 1998-04-14  
; NUMBER OF SEQ ID NOS: 147  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 23  
; LENGTH: 786  
; TYPE: PRT  
; ORGANISM: Mammalian (Human) SULU3  
US-09-291-417-23

Query Match 4.1%; Score 127.5; DB 11; Length 786;  
Best Local Similarity 20.1%; Pred. No. 0.032;  
Matches 142; Conservative 98; Mismatches 237; Indels 231; Gaps 33;

QY 3 NY-----PPTVAAQPTTTANPLQRHQSQRRRELPKIVETESTMDITIGOSKQPFUK 57  
DB 190 NYREEGDPTFASDPQSP--PQVSRHKSHYRNREHPATIRTASL-----VTRQWQHE 240  
QY 58 SIDELAAFSVAVETFKQDFDLOKHIESIENALDSKLENGVVLAAARNNHQPMSPSPR 117  
DB 241 QDSELREQMSGYKMRROH---QKQMTLENKLUKAEWDEHRLRL-----DKOLETOR 289  
QY 118 NNVSVEITTVTSQSQEIVPPT---SNKPEGRMCELMCSKGLRKIYIYANISDQAKLME 174  
DB 290 NNFAAEWEKLIKXQAAEKEAKVMSNE-----EKKFQOHI-----QAQKKE 332  
QY 175 IPSAL-----KLAKPAKFVLDCTGKFLYQORRAFTKESPMSSARQVSLILLESFLM 227  
DB 333 LNSPESQKREYKLRKEQLKEELN-----ENQSTPKKEKEWL----- 370  
QY 228 PDRGKGVKTESWKDB-----AETAAVAWKRLM-----TEGGLAAAE----- 266  
DB 371 ---SKQENICHQAEAEANLRRQYLELCRRFKRRLMLGRHNLQODLYREELNKEQ 427  
QY 267 -KMDARGLLILVACFGVPS-NFRSTDL-----DLIRMSGNEIAGALKRQFLVPMVSG 319  
DB 428 TQKDLHAMLRLRHESMQELEFRHLNTIQMRCELIRLOHOTELTNQLEYNKR----- 481  
QY 320 IVESIKRGHIEALEMVTTFQMEKFAALVLTSLKMSKESFERAKKAQSPLA----- 375  
DB 482 --ERELRR-KHV-----NEVQOQPSKLSKSLQIKKQFODTCKTQTRQYKALRNH 528  
QY 376 -----FKEAATKQALVLS-----SVNQCMEH-----KLDPAKELPGWQIKE 412  
DB 529 LLETTPKSEHKAUVRUKKEEQTAKLAILAEQYDHSINEMLSQALRLDEAQAECQVLKV 588

QY 413 QI-----VSLEKDTLQDKEMEKEKARSLSL-----MEEAALAKRMVNOQ 451  
Db 589 QIQOELELLNAYQSKIKMQAEEA-QHDELRLEQVSLRRALLEQKIEEMLA--LQNER 645  
QY 452 IKRPLSPNEMPPVTSSSYPIYDRSPSQDDDDQDEISALVSSYLGPSTSPHRSRS 511  
Db 646 TERIR-----SLLERQAREIEAFDESBRMLGFSNNVLNLS 681  
QY 512 PEYMVPLPHGLGRSVYAYEHLAPN-----SYSPGHG---HRLHRQYSPSLVHGQHPLOY 564  
Db 682 PE-----AFSHSYPGASGWSHNPTGPGPHGHPMGPPQAWG--HPMQG 724  
QY 565 SP-----PHGQQQLPYGIQVYRHSPEERYL-----GLSNQSPRSNS 604  
Db 725 GPQPMGHPGPMQGVPRGSSMGVNSPOLARTASGGRTGQMSRSTS 772  
RESULT 5  
US-09-291-417-31  
; Sequence 31, Application US/09291417A  
; Publication No. US20030050230A1  
; GENERAL INFORMATION:  
; APPLICANT: PLOWMAN, GREGORY  
; APPLICANT: MARTINEZ, RICARDO  
; APPLICANT: WYTHE, DAVID  
; TITLE OF INVENTION: STE20-RELATED PROTEIN KINASES  
; FILE REFERENCE: 240/300  
; CURRENT APPLICATION NUMBER: US/09/291,417A  
; CURRENT FILING DATE: 1999-04-13  
; EARLIER APPLICATION NUMBER: US 60/081,784  
; EARLIER FILING DATE: 1998-04-14  
; NUMBER OF SEQ ID NOS: 147  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 31  
; LENGTH: 1001  
; TYPE: PRT  
; ORGANISM: Mammalian (murine/human) SULU3  
US-09-291-417-31

Query Match 4.1%; Score 127.5; DB 11; Length 1001;  
Best Local Similarity 20.1%; Pred. No. 0.047;  
Matches 142; Conservative 98; Mismatches 237; Indels 231; Gaps 33;  
QY 3 NY-----PPTVAAQPTTANPLQHQSQRRRLPKXIVETESTSMDITIGSQKQPFLLK 57  
Db 405 NYREGDPRTASDQSP--PQVSRHKSHYRNEHFATIRTAFL-----VTRQMQEHE 455  
QY 58 SIDELAASFVAVETFKQFDDQKHTESENALDSKESNGVVLAAARNNPHQPMLSPPR 117  
Db 456 QDSELREQMSGYKRMRRQH--QKQMTUENKKAEMDEHRLRL-----DKDLETQR 504  
QY 118 NVSVETTVTSQPSOEIVPET---SNKPEGGMCMELCMCKGLRKYIYANISDQAKLME 174  
Db 505 NNFAAEWEKLIKQQAAMEKEAKVMSNE-----EKKFQOHI-----QAQKKKE 547  
QY 175 IPSAL-----KLAKPAKFVLDICGKYLOGRAFTKESPMSSARQVSLILLESFLM 227  
Db 548 LNSFLESQREYKLRKEQLKBEIN-----ENQSTPKKEQEWL----- 585  
QY 228 PDRGKGKVKIESIKDE-----AETAAVAWRKELM-----TEGGLAAAE- 266  
Db 586 ---SKQENIQHFQABEEANLLRRQCYLECRFRKRMGLGRNLEQDLVREELNKRQ 642  
QY 267 -KYDAGLLLLVACGVPFS-NFRSTDLL-----DLIRMSGNEIAGALKRSQFLVPMVSG 319  
Db 643 TQKDLBHAMLLRQHSQMQLLEFRLNTIQKRCELIRLQHOTELTNQLEYNKR- 696  
QY 320 IVESSTIKRGMHIEALEMVYTFGMEDFSAALVLTSLFKMSKESFERAKKQASPLA---- 375  
Db 697 --FRELIR-KHV-----NEVQQPKSLSKELQIKKQFQDTCKIQTOYKALENH 743  
QY 376 -----FKEAATKQALAVLS-----SVNQCMETH--KLDPAXKELPGWQIKE 412

Db 744 LLETPKSEKAVLKRLKEBQTRKLAIAEQYHSINEMLSQALRLDEQAQAEQVLEK 803  
QY 413 QI-----VSLEKDTLQDKEMEKEKARSLSL-----MEEAALAKRMVNOQ 451  
Db 804 QIQOELELLNAYQSKIKMQAEEA-QHDELRLEQVSLRRALLEQKIEEMLA--LQNER 860  
QY 452 IKRPLSPNEMPPVTSSSYPIYDRSPSQDDDDQDEISALVSSYLGPSTSPHRSRS 511  
Db 861 TERIR-----SLLERQAREIEAFDESBRMLGFSNNVLNLS 896  
QY 512 PEYMVPLPHGLGRSVYAYEHLAPN-----SYSPGHG---HRLHRQYSPSLVHGQHPLOY 564  
Db 897 PE-----AFSHSYPGASGWSHNPTGPGPHGHPMGPPQAWG--HPMQG 939  
QY 565 SP-----PHGQQQLPYGIQVYRHSPEERYL-----GLSNQSPRSNS 604  
Db 940 GPQPMGHPGPMQGVPRGSSMGVNSPOLARTASGGRTGQMSRSTS 987  
RESULT 6  
US-10-177-293-254  
; Sequence 254, Application US/10177293  
; Publication No. US20030124129A1  
; GENERAL INFORMATION:  
; APPLICANT: Lillie, James  
; APPLICANT: Glatt, Karen  
; APPLICANT: Zhao, Xumei  
; APPLICANT: Gannavarpu, Manjula  
; APPLICANT: Kamatkar, Shubhangi  
; APPLICANT: Mertens, Maureen  
; APPLICANT: Myer, Vic  
; APPLICANT: Wang, Youzhen  
; APPLICANT: Xu, Yongyao  
; APPLICANT: Hoersch, Sebastian  
; APPLICANT: Monahan, John  
; APPLICANT: Meyers, Rachel E.  
; APPLICANT: Bast Jr., Robert C.  
; APPLICANT: Hortobagyi, Gabriel N.  
; APPLICANT: Pusztai, Lajos  
; APPLICANT: Meric, Funda  
; APPLICANT: Sahin, Aysegul  
; APPLICANT: Miller, Gordon B.  
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,  
; FILE REFERENCE: MRI-038  
; TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER  
; CURRENT APPLICATION NUMBER: US/10/177,293  
; CURRENT FILING DATE: 2002-06-21  
; PRIOR APPLICATION NUMBER: US 60/299,887  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 60/301,572  
; PRIOR FILING DATE: 2001-06-27  
; PRIOR APPLICATION NUMBER: US 60/306,501  
; PRIOR FILING DATE: 2001-07-18  
; PRIOR APPLICATION NUMBER: US 60/325,002  
; PRIOR FILING DATE: 2001-09-25  
; PRIOR APPLICATION NUMBER: US 60/362,585  
; PRIOR FILING DATE: 2002-03-05  
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx  
; PRIOR FILING DATE: 2002-05-14  
; NUMBER OF SEQ ID NOS: 506  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 254  
; LENGTH: 1005  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-177-293-254  
Query Match 4.1%; Score 127.5; DB 15; Length 1005;  
Best Local Similarity 20.1%; Pred. No. 0.047;  
Matches 142; Conservative 98; Mismatches 237; Indels 231; Gaps 33;  
QY 3 NY-----PPTVAAQPTTANPLQHQSQRRRLPKXIVETESTSMDITIGSQKQPFLLK 57

Db 409 NYREEDPRTRADPQSP--PQVSRHSHYRNREHFATIRTASL-----VTRQMOEHE 459  
Qy 58 SIDELAAPSVAVETFKRODDLOKHIESIENAIKSLKESNGVVLAAARNNNHQPWLSPPR 117  
Db 460 QDELEQWGYKRMRRQH--OKQUMTLENKKAEMDEHRURL-----DKDLETOR 508  
Qy 118 NNYSVETTTVSQPSQEIIVPET---SNKPEGGRMCELMCKSLRKYIYANISDQAKLME 174  
Db 509 NNPAEAMEKLIKHOAAMEKEAKVMSNE-----EKKFQCHI-----QAQOKKE 551  
Qy 175 IPSAL-----KLAPEKAFVLDICGKFLYQGRRAFTKESPMSSARQVLSLILLESFLM 227  
Db 552 LNSFLESQKREYKLRKEQLKELN-----ENQSTPKKEKEWL-----589  
Qy 228 PDSGKGVKTESWIKDE-----AETAAVAWRKELM-----TEGGAAAB-----266  
Db 590 ---SKQENICHQFAEBEENALLRRQRYLELECRFKRMLLGRHNLQDLVRELNKRQ 646  
Qy 267 -KMDARGLLLVACFGVPS-NFRSTDLL-----DLIRMSGSEIAGALKRQFLVPMVSG 319  
Db 647 TQXDLPHAMLLRQHESMQELEPHNLTIQKRCLEIRLOHOTELTNQLEYNKR-----700  
Qy 320 IVBSSIKRGMHIBALEMVVTFGMEDKFSAAVLVTSFLKMSKESFERAKRKAQSPILA-----375  
Db 701 --BRELR-KHV-----NEVROQPKSLSKELQIKQFQDTCKIOTROYKALRNH 747  
Qy 376 -----FKBAATKQLAVLS-----SYMOCMETH--KLDPKELPGWQIKE 412  
Db 748 LLETTPKSEHKAVALKRLKEQTRKLAIAEAQDHSINEMLSLQALRLDEAQAECQVLKM 807  
Qy 413 QI-----VSLEKDTLQDKEMEERKARSL-----MEEAALAKEMVNOQ 451  
Db 808 QLOQELELLNAYOSKIKMAEA--QHDRELELEORVSLRALLLEOKIEBMLA--LQNER 864  
Qy 452 IKRPRLSPEMPPVTSYSSPIYDRSPFSDQDDQDEISALVSSYLGPESTSPFHRSSRS 511  
Db 865 TERIR-----SLLERQAREIEAEDSSMRLGFSNNVLSNLS 900  
Qy 512 PEYMWVPLPHGGLGRSVYAYEHLAPN---SYSPGHG---HRLHROYSPSLVHGQHPLOY 564  
Db 901 PE-----AFSHSYFGASGWSHNPATGGPGPHWHPGPPQAWG--HPMQG 943  
Qy 565 SP-----PIHQOOLPYGIQVRYRHSPEERYL---GLSNQRSPRSNS 604  
Db 944 GPQPGWHPSPGMQGVPRGSSMGVNSPQALRRTASGRTEQGMSESTS 991

## RESULT 7

US-09-919-172-98  
; Sequence 98, Application US/09919172  
; Patent No. US20020119463A1  
; GENERAL INFORMATION:  
; APPLICANT: Paris, Mary  
; TITLE OF INVENTION: PROSTATE CANCER MARKERS  
; FILE REFERENCE: PA-0036 US  
; CURRENT APPLICATION NUMBER: US/09/919,172  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 102  
; SOFTWARE: PERL Program  
; SEQ ID NO 98  
; LENGTH: 3256  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20020119463A1 2700132CD1  
US-09-919-172-98

Query Match

4.1%; Score 127.5; DB 10; Length 3256;

Best Local Similarity 20.0%; Pred. No. 0.31;  
Matches 137; Conservative 92; Mismatches 228; Indels 229; Gaps 32;  
Qy 11 QTTTANPLLQHQHQEQRR--RELPKIVETESTSMDI-----TIGSQKQPOFLK 57  
Db 864 ESKTVSTVNRGREGTEPRNTQKLPVESKSEETTEIVECILKRGQKATLLQQRREEMK 923  
Qy 58 SIDELAAPSVAVETFKRODDLOKHIESIENAIKSLKESNGVVLAAARNNNHQPWLS 114  
Db 924 EIE-----RPFETVKENIELKEN--DEKMK-----MKSRWTGOKCAPM-- 961  
Qy 115 PRNNVSVEITTVSQPSQEIIVPETSN-----KPEGGRMCELMCKSGLRKY 160  
Db 962 -----SDLTDLKSLPDTLMKDTARQONLLOQDHAKAPKSEKGIKTKMPCOSLOPEP 1014  
Qy 161 IVANISDQAKLMEEPKSAKLAKEPAKVLDCIGKF-YLQGRRAFTKESPMSSARQVLSL 219  
Db 1015 INTPTHTKQQL-----KASLGKGVKVELLAVGKFTRTSGTETHREHPAGDCKSI-- 1065  
Qy 220 ILESFLLMPD-----RGKGVKIESWIKDEAETAA--VAVRKELMTTEGGLAABKMDAR 271  
Db 1066 --RTFKESFKQILDPAARVTGMKWPRTPKKEAQSLDLAGFKELFQTPGPSEESMTDEK 1123  
Qy 272 GLLLLVACFGVP-----SNFRSTDL-----LDLIRMSGSEIAGALKRSQF 312  
Db 1124 --TTKIACKSPPEVSVDTPSTKQWPKESLRKADVEBEFLALRLTDS---AG---KAM 1174  
Qy 313 LPMVSGIVSESIK--RGMHIEALEMVVTFGMEDKFSAAVLVTSFLKMSKESFERAKRKA 370  
Db 1175 LTPKAGGDEKDIKAFMGTPVQKLDLACT-----LFGSKRQLQTPKEKA 1218  
Qy 371 QS--PLA-FKE-----AATKQ-----383  
Db 1219 QALEDLAGFKELFQTPGHTEBLVAAKTKTKPCSPQSDPDVDTSTKQRPKRIRKADV 1278  
Qy 384 ----LA-----VLSSVMQCMET-----HKLDPKELPGWQIKEQIVSL 417  
Db 1279 EGBELLACRLMPSAGKAMHTPKPSVGEBKDIIIFVGTVPVQKLDLTENLTGSKRRPQTPKE 1338  
Qy 418 EXDTQLQDKEMEERKARSLSLMEEAALAKEMVNOQIKRPLSPMEMPPVTSYSSPIYDR 477  
Db 1339 EQALEDLTGFKELFQTPGHTEEAVAAG-----KTKMPCESPPESADTPTSTRQRP 1391  
Qy 478 SPFSQRDDQDEISAL--VSSYLGPESTSPFHRSSPEYVMVPLPHGGLGRSVYAYEHLAP 535  
Db 1392 KTFLEKRDVQKELSALKLTQTSGETT---HTDK-----VP---GGEDKSINAFRETAK 1439  
Qy 536 NSYSPGHGRLHROYSPSLVHGQHP 561  
Db 1440 QKLDP-----AASVTGSKRHP 1455

RESULT 8  
US-09-919-039-21  
; Sequence 21, Application US/09919039  
; Publication No. US20030108871A1  
; GENERAL INFORMATION:  
; APPLICANT: Kaser, Matthew R.  
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES  
; FILE REFERENCE: PA-0035 US  
; CURRENT APPLICATION NUMBER: US/09/919,039  
; PRIOR FILING DATE: 2002-09-09  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 401  
; SOFTWARE: PERL Program  
; SEQ ID NO 21  
; LENGTH: 3256  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20030108871A1 2700132CD1

US-09-919-039-21

Query Match 4.1%; Score 127.5; DB 11; Length 3256;  
Best Local Similarity 20.0%; Pred. No. 0.31;  
Matches 137; Conservative 92; Mismatches 228; Indels 229;

QY	11	OPTTANPLLQRHQBQR--RELPIKIVETESTSMDI-----TIQSQQPQFLX	57
Ddb	864	EP\$KT\$VTVNRSGSTBFBNIOKLIPVESKEEINTBEICILUKROKATLIQORREMK	923
QY	58	SIDELAAFSVAVETFKRQFDLOKHIESIENATDSKLESNGVVLAARNNFHQ---PMLS	114
Dd	924	EIE-----RPPTYKENIELKEN-DEOKKA-----MKRSRTWGQKCAPM--	961
QY	115	PPRNNVSVETTIVVSQSQSIVPETSN-----KPEGRMCWELCMCKGLRKY	160
Dd	962	-----SDLTDLXSLEDTBELMDXTARGONLLQTODPHAKAPSEKGKITKMPCOSLOPEP	1014
QY	161	IYANISDQAKLMBEIP\$ALKLAPKAFVLVDICIGF-YLQGRRATKESPMSSARQVSLL	219
Dd	1015	INTPTHYKQOI-----KASLGKVGVKEALLAVGKETRTSGTTHREPAGDGSI---	1065
QY	220	ILE\$FLMLPD-----RGKGVKVIBSWIKDBAETA-A-VAMPKRLMTGGJLAAAEKDAR	271
Dd	1066	--RTFKESP\$KQLDPAAARVTGMKKWPRTPKEEAQSLIEDLAGFKELFOTPGPSEESMTDK	1123
QY	272	GILLLVACFGVP-----SNPRSTD-----LDLRMSGNETAGALKRSQP	312
Dd	1124	-TTKIACKSPPSVDTPTSTQWPKRSURKADVEEBFLARKUTPS---AG-----KAM	1174
QY	313	LVPVMVGIVES\$IK-RGMHEIALEMVYTFCMBDFSAALVLT\$FLMKSKSFERAKKA	370
Dd	1175	LTPKPAGGDEKDIKAPMGTPVQKLDLAGT-----LP\$KSRQLOTPEKA	1218
QY	371	QS--PLA-FKE-----AATKO-----	383
Dd	1219	QALEDLAGFKELFOTPGHTHEELVAAGKTKIPCDSQSDPVDTPTSTQWRKRSIRKADV	1278
QY	384	----LA---VLSSVMQCWET-----HKLDPAKELFGWQIKEQIVSL	417
Dd	1279	EGELLACENTLMP\$AKAMHTPKPSVGEEDKIIFVGTPTVQKLDLTENLTGSKRRPQTPKE	1338
QY	418	EKDITOLDKEMEEKARSLSLMEBAALA\$KRWYNOQIKRPLRSMWEHPPTVISSSYPIHDR	477
Dd	1339	EAQALEDTGFKELFOTPGHTEEVAAG-----KTKMPCSE\$PP\$ADTPTSTRQOP	1391
QY	478	\$FPSORDDDDEISAL--V\$SYLGPTS\$PFHR\$RR\$PEYMWPLPHGGLGR\$VYAVEHLAP	535
Dd	1392	KTFLEKRQVQKELSALKILTQTSGETT---HTDK-----VP---GGEDKSINARETAK	1439
QY	536	NSY\$PGHGHLRHROY\$PSL\$VHGQRHP	561
Dd	1440	OKLDP-----AASVTGSKEHP	1455

## RESULT 9

US-10-097-340-43  
; Sequence 43, Application US/10097340  
; Publication No. US20030087250A1

; GENERAL INFORMATION:

APPLICANT: John Monahan

APPLICANT: Manjula GANNAVARAPU

APPLICANT: Sebastian Hoersch

APPLICANT: Shubhangi KAMATKAR

APPLICANT: Steve G. Kovats

APPLICANT: Rachel E. MEYERS

APPLICANT: Michael MORRISEY

APPLICANT: Peter OLANDT

APPLICANT: AMI SEN

APPLICANT: Peter VEIBY

APPLICANT: Gordon B. Mills

APPLICANT: ROBERT C. BAST, JR.

APPLICANT: Karen LU

OF THE  
INVESTIGATION

```

/ APPLICANT: Rosemarie SCHMANDT
/ APPLICANT: Xumei ZHAO
/ APPLICANT: Karen GLATT
/ TITLE OF INVENTION: Nucleic Acid Molecules
/ TITLE OF INVENTION: Assessment, Prevention
/ FILE REFERENCE: MRI-030
/ CURRENT APPLICATION NUMBER: US/10/097,340
/ CURRENT FILING DATE: 2002-03-14
/ PRIOR APPLICATION NUMBER: 60/276,025
/ PRIOR FILING DATE: 2001-03-14
/ PRIOR APPLICATION NUMBER: 60/325,149
/ PRIOR FILING DATE: 2001-09-26
/ PRIOR APPLICATION NUMBER: 60/276,026
/ PRIOR FILING DATE: 2001-03-14
/ PRIOR APPLICATION NUMBER: 60/324,967
/ PRIOR FILING DATE: 2001/09/26
/ PRIOR APPLICATION NUMBER: 60/311,732
/ PRIOR FILING DATE: 2001-08-10
/ PRIOR APPLICATION NUMBER: 60/325,102
/ PRIOR FILING DATE: 2001-09-26
/ PRIOR APPLICATION NUMBER: 60/323,580
/ PRIOR FILING DATE: 2001-09-19
/ NUMBER OF SEQ ID NOS: 363
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 43
/ LENGTH: 1203
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ US-10-097-340-43

```

Query Match 4.0%; Score 124.5; DB 15; Length 1203;  
Best Local Similarity 19.9%; Pred. No. 0.12;  
Matches 111. Conservative 91; Mismatches 213; Indels 143;

```

QY      7 TVAAOPTTANPLQRHQS-----PQRR-----ELKIV-----36
Db     169 TIDTAPLSSVSLINKFDSQLGGQARGTGRRTMLPPQRRKSKSLDSRLPRDTEERE 228
QY     37 -----ETESTSMITICQSQKQFQKSIDELAASFVAVETFKRQDD--LQKHIESIEN 88
Db    229 RQSTNHTSSITKYDNHVGTQKQPAQSQNSPLSGFSRS-----RQTDWVLQSFEPERS 283
QY     89 AID-SKLESNGVILAEANNHPQMLSPNNVSVETTV--TVSQSQSIVETSNKPSG 145
Db    284 AQDPTMLQFKSTPDLKD---QOEAAPOGSDHMKATYIGILREGSSE--SETSVRRKV 337
QY    146 GNMCE-----LMCSKGLRKIYA-----NISD-QAKLMBEI-----PSALKLAKEPA 186
Db    338 SLVLEKMQPLVMWSSGSTKAVAGQELTRKVEBLQRLDBEVKROKLEPQSOVGLERQLE 397
QY    187 KEVLDCIGKFKYLGCRRAFTKESPMSSARQVSLILBSFULLMPDRGKGKVIKSIWKDBAE 246
Db    398 EXTBECSBLQELLERR---KGEAQOSNKE-----LQNMKLLDQGG-----LRHGLE 442
QY    247 TAAVAWRRLMTEGGLAAAKMDARGLLLVACFGVPSNFRSTDLLDLIRMGSGSNEIAGA 306
Db    443 TQVMELQNLKHVQGPPEAKEVILKDLL-----ETRELLEEV-LEQKORVEBQ 489
QY    307 LK-RQCFLVPMVSGIVESSIKRGMHIEALEWYTFGMEDKFSAAVLVTSFLXMSKESPER 365
Db    490 LRIRERETALKGALKKEVASRQDEVEHVQOY-----QBDTEQLER 531
QY    366 AKRQASQPLAFKEATKQLAVLSSWMOCMETHKLDPAKELPG-WQ-----IKE 412
Db    532 SMDQATQDHAIVLEAERGKMSALVEGLQ---RELEETSEETGHQWFMFOKNEDIRATQ 587
QY    413 QIVLSLEKDTLOLDXEMEEKARSLSMEEAALAKRWYNQOI---KRPLSPME-----MPP 464
Db    588 ELUQLRMKEKEWESELEKTEVLQRELEQBARAGSAGTRQVEVLKXELLTQBELKELQAE 647
QY    465 VTSSYSPIYDRSPPSQ 482
Db    648 ROSQEVAGRHRRDRELEKQ 665

```

```

; FILE REFERENCE: GC394C1-US
; CURRENT APPLICATION NUMBER: US/09/932.183A
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 09/308,375
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/US98/18928
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: EP9719636.4
; PRIOR FILING DATE: 1997-09-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2285
; TYPE: PR1
; ORGANISM: Bacillus subtilis
; US-09-932-183A-2

```

```

Query Match      3.9%; Score 122.5; DB 10; Length 2285;
Best Local Similarity 19.3%; Pred. No. 0.49;
Matches 83; Conservative 70; Mismatches 15; Indels 125; Gaps 15;

QY 159 KTYANISD---QAKMEEIPSAKLAKPAKFLVDCIGKIFYLOGRRRAFTKESPMSSARQ 215
Db 526 KTIIFARIGNQSSIKALEQIGISVKTAGEAKSADLISE-----VAGKWDTLSDAQK 578
QY 216 VSLILLESFLMPDRGKGVKTESWIKDEAETAA-----VAVRKRL----- 256
Db 579 QNTSIGVAGIYQLSRFNAMNMFIAQNAAKTAANSTGSAWSEQOQYADSLQARVKNLQN 638
QY 257 -MTEGGLAAAEKMDARGLLLLVACFGVPSNFRSTDLDLIRMSSG--NEIAGALKRSOFL 313
Db 639 NTFEFAIADSA-----FISDGLIETQAGSLLNASGVKISGVGL 680
QY 314 VPMVSGI-----VSSIKRGMHIEALEMVYTFGMEDKFSALV----- 351
Db 681 PPLAAVSTATLLSKNTRTLASSLILGTFRAMQETLATAGLEAGMTRAAVASRVLKTLA 740
QY 352 -----LTSP--LKNKSESEFEAKR-KAOSPLAFKEATKOLA 385
Db 741 RGLVSTLVGGAPALGWALESLSISFAEAKKAKODFQSQQTNEAITTKDSTDKLIQ 800
QY 386 VLSSVMQCMETHKLDPAKELPGWKEQIV-----SLKXDTLQDKEMEE 430
Db 801 QYKELQVKRESRLTSDSEQEVLTQQLAQTFPALVKYDQSGNAILTKNKELEKAIEN 860
QY 431 KARSLSI--MEEAALAKRMVNOQIKRPLSPNEMPPVTSYSSSPI--YDRSFPDQ---R 483
Db 861 TKYELALKQETRDSAKTTFEDASKEIKSKDEL-----KQYKQIADYNDKGRPKWDLIA 915
QY 484 DDDQDEISA 492
Db 916 DDDYKVA 924

```

```

RESULT 12
US-10-128-714-3240
; Sequence 3240, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Hu, Wenqi
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066

```

```

; FILE REFERENCE: GC394C1-US
; CURRENT APPLICATION NUMBER: US/09/932.183A
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 09/308,375
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/US98/18928
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: EP9719636.4
; PRIOR FILING DATE: 1997-09-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2285
; TYPE: PR1
; ORGANISM: Bacillus subtilis
; US-09-932-183A-2

```

```

Query Match      4.0%; Score 124; DB 15; Length 426;
Best Local Similarity 22.1%; Pred. No. 0.024;
Matches 95; Conservative 44; Mismatches 155; Indels 136; Gaps 20;

QY 173 BEIPSALKLA-KEPAKFLVDCIGK-----YLQRRAT--KESPMSSARQVSLILLES 223
Db 106 EEALDALARHEEGARIVGLCTGAFVLAAGLLDGRPATTHMYATLAKRPSVHVDP 165
QY 224 FLMPDRGKGVKTESWIKDEAETAAVAVRKRLMTEGGLAAAEKMDARGLLLLVACFGVP 283
Db 166 ELFDVDD---GDV-----LTSAGTAA----- 182
QY 284 SNFRSTDL-LDLIRMSGNEIAGALKRSOFLVPMVSG-----IVESSIKRGMHIEALEMVY 338
Db 183 ----GIDCLHIVTRDGNFAAGLARRLVVPRRSGGOERYLDRLPBEIGADPLAEV 238
QY 339 TFGME--DKPSAALVTSFLKMSKESFERAKR--AQSPPLAFKEATKQLAVLSSVMQC 393
Db 239 AWALEHLEHQFDVE--TLAARAYMSRTFDRFRSLTGSAPLQ----- 279
QY 394 METHKLDPAKELPGWKEQIVSLEKDTLQDKEMEKARSLSLMEBAALAKRMVNOQIK 453
Db 280 -----WLITQVLOQRILLETSDYSVDEVAGRCGRFSFVALRGHFRROLGS 325
QY 454 RPLSPKMPFPVTSYSSPIYDRSFPDQDDQDEISALVSLVGLPSTSFPHRSRRSPE 513
Db 326 SP-----AAVRAAYRARPQGERPVDP-----SVVGPPPALPQ-----E 360
QY 514 YMVPLPHGGLGRSVAYEHLAPN---SYSPG-HG-HPLHRQYSPSLVHGQRHPLQYS---- 565
Db 361 APVPNQ----TERTAAASSILGPAASLSTEPCKGKSHGSHGSHGSHGSHGSHGSELYAAGR 416
QY 566 PPIHQQQQLP 575
Db 417 PSLPGQRSAP 426

```

```

RESULT 11
US-09-932-183A-2
; Sequence 2, Application US/09932183A
; Patent No. US20020127641A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms

```

```

; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 3240
; LENGTH: 1001
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
; US-10-128-714-3240

Query Match      3.7%; Score 115.5; DB 15; Length 1001;
Best Local Similarity 19.1%; Pred. No. 0.54;
Matches 113; Conservative 103; Mismatches 254; Indels 123; Gaps 24;

QY      7 TVAAOPTTANPLLOHQSEQ--RRRLPKI-----VET-----ESTSMITIG 48
DB      93 TASTEITASKKELPAKESEEFSDNELPRLSELKEKEQEVETLKSOVESLKDRLAVA 152
QY      49 QSKQPOFLKSIDELAAPSVAV-----ETFKQFDLOLKHIESIENAIK 93
DB      153 RESTEGWQNLSEATRELVELDIRDKQDSEIKKLDTRQAEVDDVRAKLAESNTIATA 212
QY      94 LESNGVVLAAANNHQPMLSPRNNVSVETTVVSQPSQEIIVPETSNNKPEGGRMCLMC 153
DB      213 GEAELKELKQKTEE--ISKLOQVQCKNDNDQKQELQKLEVEAKERDASEK----- 265
QY      154 SKGLRYIYANISDQAKLMEEIPSALKAKPAKPVLDICIKF-----YLOGR----- 201
DB      266 KLGVLQGLVDNLRSLQKDTVEVVVDL---KADMKGVBELGKLQNVVDYLNNLKDNAEV 322
QY      202 -----PAFTKESPMSSARQVSLILLESFLMPDRGKGVKIESWIKDEAEV---AAVAVR 253
DB      323 QOTRDOAAGSQSPDFSLRKSAPLQO-----EPGGGNLATEPEPPANGATGGAGAKKK 377
QY      254 KRLMTGGGLAAAEKMDARGLLIL--VACFGVPSNFRSTDLDD-----LIRMGSGNE-- 302
DB      378 KNKKKGGKGEDTAKATGAAVQDQKSHAESDQAASDLADLEQIKIQLTKQLGDKAEA 437
QY      303 ---IAGALKRSQFLVPMVSGIVESSIKRGM-HIEALEMVIYFGMEDKF-----SA 348
DB      438 IDRLSKLKGEEGLKEBIEISLDDLLHGDHVEAKOKIKELNVKKALETVSKLEKEL 497
QY      349 ALVLTFLKMS-----KESFERAKRKAQSPLAFKAEATKQAV--LSSVMQCMET 396
DB      498 ADIRTSKASKSADSEKMHSDLKEDYENLVKLTN-LETLSAAQQLAATRFKDLTELRET 556
QY      397 -HKLDPAKELPGHQI-----KEQIVSLEKDTQLDKEMEE-----KARSLSLMEEAA 442
DB      557 LQKLQP--ELKSURVSESELKSTKEALASKESBELRTLEGKHEELAEVTKLSTISERDA 614
QY      443 LAKRWYNOQIKRPLSPMEMPPTVSSSYPIYDRSFPSQRDDDDQDEISALVS 495
DB      615 EVKTL-NQKIRQETDSRLKABENLITVAQSDRLVSESKQEAETVEKIAADLS 666

RESULT 13
US-10-177-293-391
; Sequence 391, Application US/10177293
; Publication No. US20030124129A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavarpu, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen

```

```
US-10-171-311-164
; Sequence 164, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersb, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 164
; LENGTH: 1938
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-164

Query Match 3.7%; Score 114.5; DB 15; Length 1938;
Best Local Similarity 19.7%; Pred. No. 1.9;
Matches 124; Conservative 90; Mismatches 205; Indels 211; Gaps 26;

Qy 22 RHQSEORRELPKIVETESTSMDITIGSQKQPOFLKSIDELAAFSVAVETFKRQFDD--- 78
Db 1134 RNKAEKQKRDGELEBALKTELEDTLDSTATQOELRAKRE-----QEVTVLKALDEETR 1188
Qy 79 -----LQKHIESIENAIKLENGVVLAAARNNHQPMLSPPRNVSVEITTVTS 129
Db 1189 SHEAQVOEMRQKHAQAVEE-LTEQLEQ--FKRAKANLKNKOTLEKENADLAGELRV-LG 1244
Qy 130 QPSOEIVPETSNTKEGGRMCELM--CSKGLRKYYIANISDQA-KLMBEIPS----- 177
Db 1245 QAKQEV--EHKKKKLEAQVQLQSKCSDGER--ARAELNDKVHKLQNEVESVTGMLNEAE 1300
Qy 178 --ALKLAKEPAKF-----VLDCIGKPYLQGRRAFTK-----ESPM 210
Db 1301 GKAIKLAQVSLSSQLQDTQELLQEBTRQKLVSTKQLQLEBEERNSLQDQLEMEAKQ 1360
Qy 211 SSARQVSLILLESFLMPDRGK-----GKVKIESWIKD---EAETAAVAMR 253
Db 1361 NLERHISTINIQ-----LSDSKKLQDFASTVEALEBEGKRFQKEIENLTQOYEKAAAYD 1416
Qy 254 KRLMTEGGL-----AAAEKMDARGILLVLVACFGVPSNF----- 286
Db 1417 KLETKNLQOELDDLVDLDNQVLVSNLEKKQKQKFDQLLAEEKNISKYADDERDRAEA 1476
Qy 287 ----RSTDLLDLIR-MGSGNEIAGALKR-SQFLVPVMSGIVESSIKRGWHIEALE----- 335
Db 1477 EAREKETKALSARALEEALAEKEELERTNKMKAEMEDLVSSKDDVGNVHELEKSKEA 1536
Qy 336 -----MVYTFGMEDKFSAA-----LVLSFLKMSKESFER----- 365
Db 1537 LETQMEEMKQLEBELELQATEDAKLRLEVNMQALKGQFERDLQARDQNEBEKRQOLR 1596
Qy 366 -----AKRK-----AQSPFLAFKEAATKQAVLSSVMQ- 392
Db 1597 QLHEYETELEDERKORALAAAAKKKLEGLDKLQLEQADSAIKGREBAIKQLRKLAQMKD 1656
Qy 393 -----CMETHKLDPAKELPGWQIKQEVLSLEKDTLQLDKMEBEKARS 434

US-09-927-597-2
; Sequence 2, Application US/09927597
; Publication No. US20030032018A1
; GENERAL INFORMATION:
; APPLICANT: Malik, Fady
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; APPLICANT: Craven, Andrew
; APPLICANT: Sakowicz, Roman
; APPLICANT: Hartman, James
; TITLE OF INVENTION: Human smooth muscle myosin heavy chain
; FILE REFERENCE: CITOP018
; CURRENT APPLICATION NUMBER: US/09/927,597
; CURRENT FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1945
; TYPE: PRT
; ORGANISM: Human
US-09-927-597-2

Query Match 3.7%; Score 114.5; DB 11; Length 1945;
Best Local Similarity 19.7%; Pred. No. 2;
Matches 124; Conservative 90; Mismatches 205; Indels 211; Gaps 26;

Qy 22 RHQSEORRELPKIVETESTSMDITIGSQKQPOFLKSIDELAAFSVAVETFKRQFDD--- 78
Db 1141 RNKAEKQKRDGELEBALKTELEDTLDSTATQOELRAKRE-----QEVTVLKALDEETR 1195
Qy 79 -----LQKHIESIENAIKLENGVVLAAARNNHQPMLSPPRNVSVEITTVTS 129
Db 1196 SHEAQVOEMRQKHAQAVEE-LTEQLEQ--FKRAKANLKNKOTLEKENADLAGELRV-LG 1251
Qy 130 QPSOEIVPETSNTKEGGRMCELM--CSKGLRKYYIANISDQA-KLMBEIPS----- 177
Db 1252 QAKQEV--EHKKKKLEAQVQLQSKCSDGER--ARAELNDKVHKLQNEVESVTGMLNEAE 1307
Qy 178 --ALKLAKEPAKF-----VLDCIGKPYLQGRRAFTK-----ESPM 210
Db 1308 GKAIKLAQVSLSSQLQDTQELLQEBTRQKLVSTKQLQLEBEERNSLQDQLEMEAKQ 1367
Qy 211 SSARQVSLILLESFLMPDRGK-----GKVKIESWIKD---EAETAAVAMR 253
Db 1368 NLERHISTINIQ-----LSDSKKLQDFASTVEALEBEGKRFQKEIENLTQOYEKAAAYD 1423
Qy 254 KRLMTEGGL-----AAAEKMDARGILLVLVACFGVPSNF----- 286
Db 1424 KLETKNLQOELDDLVDLDNQVLVSNLEKKQKQKFDQLLAEEKNISKYADDERDRAEA 1483
Qy 287 ----RSTDLLDLIR-MGSGNEIAGALKR-SQFLVPVMSGIVESSIKRGWHIEALE----- 335
Db 1484 EAREKETKALSARALEEALAEKEELERTNKMKAEMEDLVSSKDDVGNVHELEKSKEA 1543
Qy 336 -----MVYTFGMEDKFSAA-----LVLSFLKMSKESFER----- 365
Db 1544 LETQMEEMKQLEBELELQATEDAKLRLEVNMQALKGQFERDLQARDQNEBEKRQOLR 1603
Qy 366 -----AKRK-----AQSPFLAFKEAATKQAVLSSVMQ- 392
Db 1604 QLHEYETELEDERKORALAAAAKKKLEGLDKLQLEQADSAIKGREBAIKQLRKLAQMKD 1663
Qy 393 -----CMETHKLDPAKELPGWQIKQEVLSLEKDTLQLDKMEBEKARS 434

US-09-927-597-2
; Sequence 2, Application US/09927597
; Publication No. US20030032018A1
; GENERAL INFORMATION:
; APPLICANT: Malik, Fady
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; APPLICANT: Craven, Andrew
; APPLICANT: Sakowicz, Roman
; APPLICANT: Hartman, James
; TITLE OF INVENTION: Human smooth muscle myosin heavy chain
; FILE REFERENCE: CITOP018
; CURRENT APPLICATION NUMBER: US/09/927,597
; CURRENT FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1945
; TYPE: PRT
; ORGANISM: Human
US-09-927-597-2

Query Match 3.7%; Score 114.5; DB 11; Length 1945;
Best Local Similarity 19.7%; Pred. No. 2;
Matches 124; Conservative 90; Mismatches 205; Indels 211; Gaps 26;

Qy 22 RHQSEORRELPKIVETESTSMDITIGSQKQPOFLKSIDELAAFSVAVETFKRQFDD--- 78
Db 1141 RNKAEKQKRDGELEBALKTELEDTLDSTATQOELRAKRE-----QEVTVLKALDEETR 1195
Qy 79 -----LQKHIESIENAIKLENGVVLAAARNNHQPMLSPPRNVSVEITTVTS 129
Db 1196 SHEAQVOEMRQKHAQAVEE-LTEQLEQ--FKRAKANLKNKOTLEKENADLAGELRV-LG 1251
Qy 130 QPSOEIVPETSNTKEGGRMCELM--CSKGLRKYYIANISDQA-KLMBEIPS----- 177
Db 1252 QAKQEV--EHKKKKLEAQVQLQSKCSDGER--ARAELNDKVHKLQNEVESVTGMLNEAE 1307
Qy 178 --ALKLAKEPAKF-----VLDCIGKPYLQGRRAFTK-----ESPM 210
Db 1308 GKAIKLAQVSLSSQLQDTQELLQEBTRQKLVSTKQLQLEBEERNSLQDQLEMEAKQ 1367
Qy 211 SSARQVSLILLESFLMPDRGK-----GKVKIESWIKD---EAETAAVAMR 253
Db 1368 NLERHISTINIQ-----LSDSKKLQDFASTVEALEBEGKRFQKEIENLTQOYEKAAAYD 1423
Qy 254 KRLMTEGGL-----AAAEKMDARGILLVLVACFGVPSNF----- 286
Db 1424 KLETKNLQOELDDLVDLDNQVLVSNLEKKQKQKFDQLLAEEKNISKYADDERDRAEA 1483
Qy 287 ----RSTDLLDLIR-MGSGNEIAGALKR-SQFLVPVMSGIVESSIKRGWHIEALE----- 335
Db 1484 EAREKETKALSARALEEALAEKEELERTNKMKAEMEDLVSSKDDVGNVHELEKSKEA 1543
Qy 336 -----MVYTFGMEDKFSAA-----LVLSFLKMSKESFER----- 365
Db 1544 LETQMEEMKQLEBELELQATEDAKLRLEVNMQALKGQFERDLQARDQNEBEKRQOLR 1603
Qy 366 -----AKRK-----AQSPFLAFKEAATKQAVLSSVMQ- 392
Db 1604 QLHEYETELEDERKORALAAAAKKKLEGLDKLQLEQADSAIKGREBAIKQLRKLAQMKD 1663
Qy 393 -----CMETHKLDPAKELPGWQIKQEVLSLEKDTLQLDKMEBEKARS 434
```

Db 1664 FORELEDARSRDEIFATAKENEKAKSLEADLMQLOEDLAAERARKQADLEKEELAE 1723

QY 435 LSLMEEAALAKRMVQIKR---PRLSPME 461

Db 1724 LA---SSLGSRNALQDEKRRLEARIAQLE 1749

Search completed: August 14, 2003, 10:49:26  
Job time : 62 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 14, 2003, 10:34:59 ; Search time 42 Seconds  
(without alignments)  
1394.446 Million cell updates/sec

Title: US-09-890-475-1  
Perfect score: 3104  
Sequence: 1 MSNYPPTVAAQPTTANPLL.....RYLGLSNQRSPRNSLDPK 609

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616862 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 75:\*

1: Piri:\*

2: Piri:\*

3: Piri:\*

4: Piri:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query %	Description
1	2653.5	85.5	578	2 T01229
2	2295	9.5	473	2 A86442
3	286.5	9.2	862	2 T01798
4	234.5	7.6	507	2 C71412
5	142	4.6	725	1 JC5016
6	139.5	4.5	1201	2 T08603
7	136	4.4	1722	1 I78879
8	134	4.3	841	2 A86188
9	134	4.3	2116	2 A26555
10	132	4.3	5327	2 T13564
11	131	4.2	993	2 S49461
12	129	4.2	946	2 S28061
13	127.5	4.1	2897	2 A48666
14	127.5	4.1	3256	2 A48666
15	126.5	4.1	602	2 T13988
16	126.5	4.1	1004	2 B69483
17	126.5	4.1	1920	2 A53188
18	126	4.1	694	2 T25685
19	125	4.0	1128	2 A49960
20	125	4.0	1974	2 T16703
21	124.5	4.0	1127	2 T32404
22	124.5	4.0	1538	2 T29095
23	124	4.0	3488	2 T34418
24	122.5	3.9	650	2 T00081
25	122.5	3.9	1034	2 T32297
26	122.5	3.9	2285	2 T12796
27	122	3.9	1325	2 T42722
28	121.5	3.9	279	2 D71453
29	121.5	3.9	3187	2 JC5937

serine/threonine p  
myosin-like protei  
protein F23M19.10  
hypothetical prote  
hypothetical ABC t  
protein C4IG7.5 [i  
protein K09H11.3 [i  
serine/threonine-s  
membrane associate  
still life protein  
still life protein  
hypothetical prote  
kinesin-related pr  
early endosome ant  
hypothetical prote  
hypothetical prote

ALIGNMENTS

RESULT 1

T01229  
hypothetical protein F6N23.25 - Arabidopsis thaliana  
C1Species: Arabidopsis thaliana (mouse-ear cress)  
C1Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 22-Oct-1999  
C1Accession: T01229  
R1Geisel, C.  
submitted to the EMBL Data Library, April 1998  
A1Description: The sequence of A. thaliana F6N23.  
A1Reference number: Z14281  
A1Accession: T01229  
A1Status: translated from GB/EMBL/DBJ  
A1Molecule type: DNA  
A1Residues: 1-578 <GEI>  
A1Cross-references: EMBL:AF058919; NID:G3047100; PID:G3047115; GSPDB:GNC0063; ATSP:F6N2.  
C1Genetics:  
A1Gene: ATSP:F6N23.25  
A1Map position: 5  
A1Introns: 313/3; 530/1

Query Match	85.5%;	Score	2653.5;	DB	2;	Length	578;
Best Local Similarity	83.6%;	Pred. No.	2.8e-164;				
Matches	531;	Conservative	3;	Mismatches	8;	Indels	93;
Gaps	2;						
QY	1	MSNYPPTVAAQPTTANPLLORHOSQRRRLPKIVETESTSMDITIGQSKQPOFLKSID	60				
Db	1	MSNYPPTVAAQPTTANPLLORHOSQRRRLPKIVETESTSMDITIGQSKQPOFLKSID	60				
QY	61	ELAAPSVAVEIFKRFQDDLOKHIESIENAIKLSNGVWLAARNNNFHQPMSPPPNNV	120				
Db	61	ELAAPSVAVEIFKRFQDDLOKHIESIENAIKLSNGVWLAARNNNFHQPMSPPPNNV	120				
QY	121	SVETTTVSQSQEIVPETSNNKPEGRMCELMCKGURKIYANISDAKLMEEIPSAK	180				
Db	121	SVETTTVSQSQEIVPETSNNKPEGRMCELMCKGURKIYANISDAKLMEEIPSAK	180				
QY	181	LAKEPAKFVLDICIGKFKYQGRRAFTKESPMSSARQVSLILLESFLLMPDRGKGVKIESW	240				
Db	181	LAKEPAKFVLDICIGKFKYQGRRAFTKESPMSSARQVSLILLESFLLMPDRGKGVKIESW	240				
QY	241	IDKEAFTAAVAKRLMTGEGGLAAEKDAAGLLLVACFGVPSNFRSTDLDIRMSG	300				
Db	241	IDKEAFTAAVAKRLMTGEGGLAAEKDAAGLLLVACFGVPSNFRSTDLDIRMSG	300				
QY	301	NIAGALKESQFLVPMWVGIVESSIKRGMHIEALEMVTYTFGMEDKFGAALVLTSLKMSK	360				
Db	301	NIAGALKESQFLVPMWVGIVESSIKRGMHIEALEMVTYTFGMEDKFGAALVLTSLKMSK	360				
QY	361	ESFERAKRAQSPFLAFKAAATKQLAVLSSVQCQETHKLDPAKELPGWIKQIVSLEKD	420				
Db	361	ESFERAKRAQSPFLAFKAAATKQLAVLSSVQCQETHKLDPAKELPGWIKQIVSLEKD	420				
QY	314	-----KEAATKQLAVLSSVQCQETHKLDPAKELPGWIKQIVSLEKD	357				
Db	314	-----KEAATKQLAVLSSVQCQETHKLDPAKELPGWIKQIVSLEKD	357				

QY 421 TLQDKEMEERKARSLIMEEAAIAKRYNNOQIKRPLSPNEMPPVTSSSYSPYDRSRSP 480  
DB 358 TLQDKEMEERKARSLIMEEAAIAKRYNNOQIKRPLSPNEMPPVTSSSYSPYDRSRSP 417  
QY 481 SQRDDDDQDEISALVSSYLGPGTSPPHRSRSPPEYVWPLPHGGIGRSYVAYEHLAPNSYSP 540  
DB 418 SQRDDDDQDEISALVSSYLGPGTSPPHRSRSPPEYVWPLPHGGIGRSYVAYEHLAPNSYSP 477  
QY 541 GGHRLHROYSPLVHGQPHLPYSPPIHGQQQLPYGIQRYVPHSPSEERY-----591  
DB 478 GGHRLHROYSPLVHGQPHLPYSPPIHGQQQLPYGIQRYVPHSPSEERY-----537  
QY 592 -----LGLSNQSRSPSNSS 605  
DB 538 MNVMSAFLSLRXNWSPKIQLELDNQSNPKQKQS 572

RESULT 2  
A86442  
hypothetical protein F5M6.26 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001  
C:Accession: A86442  
R:Neologs, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: A86442  
A:Status: preliminary  
A:Map position: 1-473 <STO>  
A:Residues: 1-473  
A:Cross-references: GB:AE005172; NID:g11136721; PIDN:AAG31302.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1

Query Match 9.5%; Score 295; DB 2; Length 473;  
Best Local Similarity 22.5%; Pred. No. 1.6e-11;  
Matches 124; Conservative 109; Mismatches 219; Indels 98; Gaps 21;

QY 39 ESTMDITIGSKQFQFLKSIDEIAAP-SVAVETFKQFDDLQKHIESIENAIKSLN 97  
DB 5 ESTAASINQIDKKQKLLKAFDDLOAHRSLSPFNLSWSIEDSHFSLSQSLFNRLQS- 63  
QY 98 GVVLAARNHFQPLSPRNNSVETTVTSQSQEIVPETSNNKPGGRCMLMCKSL 157  
DB 64 -----AVTSN-----SGNIETPAVTTETPVL-----WPLRFECKNDGKGL 102  
QY 158 RKYIYANISDQAKLMEIPSAKAKPAKVLDCIGKFYQGRRAFTKSPMSSAR--- 214  
DB 103 GNYMTENSRKLSINEELPNAIRCSENPAVLDAI-----EGSYHCSFSPSSSSARAID 157  
QY 215 --QVSLILLESLLMPDRGKGVKTESIKDEATAVAVAKRLMTGGLAAEKMDARG 272  
DB 158 VKRIFVILLLEAIEI-----NANUTNDRERARTIADWKPNI-----GNKPSALG 204  
QY 273 LILLVACFGVPSNFRSTDLDDLRMSGNEIAGAL-KRSQFLVPMVSGIVESSIKRGMHI 331  
DB 205 FLHLVAAPFELGSLFTEBICDYIFLSKYQATTICKIGLDNRNIGVYVQKFLDTGRLL 264  
QY 332 BALEWYITFGMDKSAALVLTSLFKMSKESFERAKRAQSPAPK-BAATKQIAVLSSV 390  
DB 265 VAIRFIYENYMGSEFEPVSILKTSIKNREAKRVCAGNYSLVKQNEATKELSAIRAV 324  
QY 391 MOCMETHKLDPAKLPGMQIKQIVLSKDTQLQDKEMEERKARSLIMEEAAIAKRYNNOQ 450  
DB 325 IKVVKEKNTES-----EFMEE--KLBEYKELEDQAKRKA-TKFNSPANPQOPQEQ 374

QY 451 QI--KRPRL---SPME-----MPVTSSSYSPYDRSRSPSQRDDDDQDEISALVSSYSP 501  
DB 375 KVNKKRPRVANGSMENYNTIPPLRQOQPPPL-----LPTPSQILQVNPYGLLSAIL-PG 428  
QY 502 TSPFHRSSRSPPEYVWPLPHGGIGRSYVAYEHLAPNSYSPGHGRLHROYSPLVHGQPH 561  
DB 429 VAVPYGNPRALFGSVAP---ASRPVYVYQ-----QTGYG-----MPP 463  
QY 562 LOYSPPIHGQ 571  
DB 464 PQYRPPYYPQ 473

RESULT 3  
T01798  
hypothetical protein A\_TM021B04.8 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 26-Feb-1999 #sequence\_revision 26-Feb-1999 #text\_change 22-Oct-1999  
C:Accession: T01798  
R:Dante, M.; Wamsley, P.; Gibson, A.  
submitted to the EMBL Data Library, June 1997  
A:Description: The sequence of A. thaliana TM021B04.  
A:Reference number: Z14440  
A:Accession: T01798  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-862 <DAN>  
A:Cross-references: EMBL:AF007271; NID:g2191181; PID:g2191193; GSPDB:GN00063; ATSP:A\_TM0  
C:Genetics:  
A:Gene: ATSP:A\_TM021B04.8  
A:Map position: 5  
A:Introns: 220/1; 280/3; 379/1; 418/1; 615/1; 674/3  
Query Match 9.2%; Score 286.5; DB 2; Length 862;  
Best Local Similarity 22.1%; Pred. No. 1.3e-10;  
Matches 132; Conservative 119; Mismatches 231; Indels 115; Gaps 21;

QY 18 PLLQHQHSEQRRELKPKIVETESTMDITIGSKQFQFLKSIDEIAAPSVAVETFKQFD 77  
DB 355 PYEQKHVKK-----PRLTPTAPSNQTLVCFNMKHELTSVNLSIF-VFGETLRRHF- 407  
QY 78 DLQKHIESIENAIKSLNNGVLAARNN-----PHQMLSPRNNSVETTVTSQ 130  
DB 408 -----NQNIFLLFKLKQPEV-----VCVPTGKQVKE 434  
QY 131 PSQIVPETSNNKPGGRCMLMCKSG-----LRKYIYANISDQAKLMEIPSAKLAKE 184  
DB 435 SGADHQPDITATHPSGTETKLNILSGSIKADMJRELVE---KQPLKESDLSNALKCTPD 491  
QY 185 PAKFVLDICIGKFYQGRRAFTKSPMSSARQVSLILLESFLMPDRGKGVKIESWIKDE 244  
DB 492 PAKFLDTSMALCPTNTEGGYEFKMLITSASCSSLNQLKLLP-----KIGHVPKGD 544  
QY 245 AETAAVAKRLMTGGLAAEKMDARGLLLVACFGVPSNFRSTDLDDLRMSGNEIA 304  
DB 545 AKLAVYWKDKI-----AKSKRDQLEVICFQIGFIVSEFKADLLGLDLSNYSWQTVS 600  
QY 305 GALKSQFLVPMVSGIVESSIKRGMHIEALEWYITFGMDKFS-AALVLTSLFKMSKESF 363  
DB 601 PDLQCFGLDDAIFGTONIKTGHRKAIKADIYISGVMVHRFQPVSAIINDSLRITKESA 660  
QY 364 BRAKRAQSPAPKAA-TKQIAVLSSVMOCMETHKLDPAKELPGMQIKQIVLSKDTIL 422  
DB 661 EKSRYEAKNSETTQVAAIDRQVRAALRAAIKICISCHKESEFQL--GDLEEQIKSLK--L 716  
QY 423 QLDKEMEERKARSLIMEEAAIAKRYNNOQIKRPLSPNEMPPVTSSSYSPYDRSRFPSPQ 482  
DB 717 RNTNSNGSGSGSASKPDSTIKQ---SQTAKPTVA--EVAPTS-----NIPLE 761  
QY 483 RDDDDQDEISALVSSYSPSTSPFHRSSRSPPEYVWPLPHGGLGRSV---YAYEHLA--PN 536  
DB 762 PSTEAASSA-----SKPFSKKNKR-----CKKRSMSGNNQSSGSHASHTS 802

QY 537 SYSPGHGRLHROYSFSLVGHQHPLOVSPPIHQQQ--LPYGIQRYVRHSPSEERY 591  
DB 803 NHYPSHDYSNQLRTPVDNYDRGTFPNDYNNQWQGPQFQYHLYQPLDPY 859

RESULT 4  
C71412  
hyaluronan receptor - human  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
A:Variety: Columbia  
C:Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 05-Dec-1998  
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk  
P.; Wedler, H.; Wedler, S.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terry, N.; Giel  
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.  
Nature 391, 485-488, 1998  
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech  
erhoft, A.; Moore, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ana  
C.; Chalwatzis, N.  
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal  
A:Reference number: A71400; MUID:98121113; PMID:9461215  
A:Accession: C71412  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-507 <BEV>  
A:Cross-references: GB:Z97337; NID:G2244829; PID:e326836; PID:G2244847  
C:Genetics:  
A:Map position: 4C09-4G3845

Query Match 7.6%; Score 234.5; DB 2; Length 507;  
Best Local Similarity 19.2%; Pred. No. 1.4e-07;  
Matches 117; Conservative 90; Mismatches 214; Indels 187; Gaps 21;

QY 49 QSKQPOFLKSIDELAAFSVAVETFKRQ-----FDDLOKHIESIENAIIDSKLESNG 98  
DB 12 KSSQSFEE-----FQQAQLMTSCNLLMKELSEHTSMEQNLMKSEALR 57

QY 99 VVLAARNNFHQPMLSPNNVSVETTVTS-----QPSQIVPETSNDK 143  
DB 58 QMIETLDNQTSIIELKREVDHSIEAGKVEERVRAALDSLEKARDCGEDTGEV 117

QY 144 EGG-----RMCELMCSKGLRKYIYANISQAKLMEIPSAKLAKPAKFLVDCIGK 195  
DB 118 DGDGGLSALKSLCKDARGFWGFIARKLEENLRQIPVALVDCVDPKLVLEAVSE 177

QY 196 FYLQGRRAFTKESPMSSARQVSLILLESF---LLMPDRGKGVKIESWIKDEAPAAV 252  
DB 178 VFPVKRG-GEKVSNDFGWACVWVLESIPVMVDPVWGKSLVTPSVYKAKAIAETW 236

QY 253 RKRLMTEGGLAAAKMDARGLLLVACFGVPSNFRSTDLDDLRMSGNEIAGALKRSQF 312  
DB 237 KASLEERGGIENKTPDVHTFLQHLVTFGIVKK---DDLALYR---KLVVGSARWQOM 280

QY 313 -----LVPMVSGIVBESIKRGMHIEALEMVYTFGMEDKFSALVLTSPFKMSKESFE 364  
DB 289 PKLAVSVGLGDQMDMEELIIRGQQLDAVHTFVGLVHLPPVPLLKAYLRDAKATA 348

QY 365 -----RAKKAQSPFLAFKAATKQLAVLSSVMQCWETHKLDPAKELPGWIKSQI 414  
DB 349 LIITDDSNNSGRSAGSVSSALAY-----QCMCEKTKTKRKA-----V 387

QY 415 VSLBKDTLQDKEMEEKARSLMEBAALAKRMYNQIKRPLSPMPPVTSSTSPY 474  
DB 388 IPANKRT-----RASYSG-----PMPKAKAGIT--- 411

QY 475 RDRSFPSQRDDQDEISALVSSYLGSTSPFPHRSRSPDEYMPPLPHGGLGRSVYAYEHLA 534  
DB 412 -----NAVVSFPPPPPTFIRSQSHSPQYGPV-----AYTTSP 444

QY 535 PNV---SPGHGHLHROYSFSLVHG--QRHPLQY-----SPPI-HGQQOL 574  
DB 445 PTIYNRSPPY-----QYSPFAVHGSYQTSYPTATYTCVSPVAAAPPPVPHPHPH 498

QY 575 PYGIQRYV 582  
DB 499 HHIIQHAY 506

RESULT 5  
JC5016  
hyaluronan receptor - human  
C:Species: Homo sapiens (man)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 29-Sep-1999  
R:Wang, C.; Entwistle, J.; Hou, G.; Li, Q.; Turley, E.A.  
Gene 174, 299-306, 1996  
A:Title: The characterization of a human RHAMM cDNA: Conservation of the hyaluronan-bin  
A:Reference number: JC5016; MUID:97045829; PMID:8890751  
A:Contents: breast  
A:Accession: JC5016  
A:Molecule type: mRNA  
A:Residues: 1-725 <WAN>  
A:Cross-references: GB:U29343  
A:Note: it is uncertain whether Met-1 or Met-196 is the initiator  
C:Comment: This receptor regulates focal adhesion turnover, and regulates human breast  
C:Genetics:  
A:Gene: GDB:HMWR; RHAMM  
A:Cross-references: GDB:683209; OMIM:600936  
A:Map position: 5q33.2-5qter  
C:Superfamily: hyaluronan receptor

Query Match 4.6%; Score 142; DB 1; Length 725;  
Best Local Similarity 20.0%; Pred. No. 0.24;  
Matches 112; Conservative 91; Mismatches 220; Indels 136; Gaps 21;

QY 34 KIVETESTMDITIGQSKQPOFLKSIDELAAFSVAVETFKRQPDLDQKHIESIENAIIDS- 92  
DB 216 KLVSTKEKID---EKSETKLELYEIEISCASQVQKYLDTAQLSEENLKEKNDEILSL 272

QY 93 --KLSSNGVLA-----ARNNFHQPMLSPNNVSVETTVTSQ 130  
DB 273 KOSLEENIVLSKQVEDLVNKKCOLLEKEKEDHVNREHNENLNAEQNL----- 322

QY 131 PQSEIVPETSNDKPEGRVCMELCSKGLRKYIYANISQAKL---MEIPSAKLAKSPAK 187  
DB 323 -KQKFILE--QSEHEKLOQKELQIDSLLOQKELSSSLHQKLCSPQSEEMVKKNLFEELK 380

QY 188 FVLDCIGKTYL---QGRRAFTKESPMSSARQVSLILLESFLLMPDRGKGVKIESWIKDE 244  
DB 381 QTLDELDKLQKKEQAERLVKOLEEAKSRAEELKLEKL-----KGG-----EAE 427

QY 245 AETAAVAMFKRLMTEGGLAAAKMDARGLLLVACFGVPSNFRS----- 288  
DB 428 LEKSSAA-----HTQATLLQEKYDS---NVQSLEDVTAQFESYKALTASEIEDLKLEN 478

QY 289 TOLLDIRMSGN-----EIAGALKRSQFLVPMVSGIVBESIKRGMHIEALEMVYTFGM 342  
DB 479 SSIQEKAAKAGKNAEDVQHQIILATSSNQEVYRMLDLQTSALKETEIKETVTSFLQKI 538

QY 343 EDKFSALVLTSPFKMSKESPERAKRKAQSPFLAFKAATKQLAVLSSVMQCWETHKLDPA 402  
DB 539 TD-----LQNLQKQBEDFRKQLEDEGRKAENKTAEI-----T 574

QY 403 KELPGHQ--IKEQIVSLEKD--TLQDK--EMEEKARSLMEBAALAKRMYN----- 449  
DB 575 EEINKVRLLYEELYNKTKPQQLDAFEVEKQA-----LLNEHGAAQEOINKIRDSYAKLL 630

QY 450 --QQIKRPLSPMPPVTSSTSPYIRDRSPSPSQDDQDEISALVSSYLG-----PST 502  
DB 631 GHQNLQKQIKHVVKLKDENSQKSEVSKLRCLQAKKQSETKLQEELNKVLGIKHFDPSK 690

QY 503 SFPHRSRSPDEYMPPLPHG 521  
DB 691 AFHESKENFALKTPKEG 709

## RESULT 6

T08603  
Kinesin-related protein COS2 - fruit fly (*Drosophila melanogaster*)  
C:Species: *Drosophila melanogaster*  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 17-Nov-2000  
C:Accession: T08603  
R:Stinson, J.C.; Ho, K.S.; Suyama, K.; Scott, M.P.  
Cell 90, 235-245, 1997  
A:Title: Costal2, a novel kinesin-related protein in the Hedgehog signaling pathway.  
A:Reference number: Z16453; MUID:97386449; PMID:9244298  
A:Accession: T08603  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1201 <SIS>  
A:Cross-references: EMBL:AF019250; NID:92330020; PIDN:AAB66813.1; PID:92330021  
C:Genetics:  
A:Gene: costal-2  
A:Cross-references: FlyBase:FBgn0000352  
A:Map position: 2  
C:Keywords: cytosol; microtubule binding; signal transduction

Query Match 4.5%; Score 139.5; DB 2; Length 1201;  
Best Local Similarity 19.7%; Pred. No. 0.74;  
Matches 132; Conservative 97; Mismatches 233; Indels 209; Gaps 29;  
Db 18 PFLQRHQSEORRELKIVETESTMDITIGSKQPOFLKSIDELAFAVAVETFKRQFD 77  
498 PLEAEPEEDVNSEANSESPNSDNEDTNSHRPDLDDKIBSL-----MEFPRKTD 551  
Db 78 D-----LOKHIESIENAI-----NG-----VVLARNNNFHQMLSPRRNN 119  
120 VSEVTTVTVSQPSQIIVPET-----SNKPEGGRCMELMCS-----KGLRKYIYA 163  
611 LAMLRNVAQQPPPIIDPESVVDPLESSSGEIGIQALAAATAPIEQOKLRKLV-A 669  
164 NTSQAKLMEEIPSAKLA-----KEPAKVLDC-----IGK 195  
670 EIEGKQRLREIETIQVKNIIAELVNSDTRSHAKQRFHKKALEACDRAKAKQLGK 729  
196 FYLQGR-----RAFTKESPM-----SSARQVSL-----ILESFLMPDRG 231  
730 ALVQGRGQSEIERWTITGHLEERLEDLSMKHTAGESGQVKLQOSVGESRKQADDLQ 789  
232 KGVKIESWIKDEASTAAVAVAKRLMTREGGLAAAE---KMDARGLLLVACFGVPSNFRS 288  
790 K-KLRKECKLRQMEAEVLKRESRETGKELVKAQSGPEQGRQKAVQA-----RI 840  
289 TDLDLIRMSGN-----ETAGALKRSQFLVPM 316  
841 TLNHLIRKSNLSEQPGPQOQETLRHEINLRGTRDLLLEERCHLDKLRKDKVLTQK 900  
317 VSGIVESSIKRGWHITALEMYVTFGMEDKFAALVLTSFLKMSKESFERAKRKAQPLAF 376  
901 EE---RKLLCEDEATEADAIEFPKNEMITGHRSDTSD-RIQREKGEQMLMARLNELST 956  
377 KBAATKQALVLSVMQC-----NETHKLDPAKLPQWQKEQIVSL-----EKOT 421  
957 EEMRTLLYKYFTKVIDLRDSRKLEQLVQLERDAMKWKERYLSNVAQARLEGHERNA 1016  
422 LQDKMEMEKARSLSIM-----EEAALAKRMYNQOIKRPRLSMPMPPTSSS---YSPI 473  
1017 VLLQRCHEMK---LTLMLRHAEETISASASGYERALP---ACVAPPVQASDDFDYDFH 1070  
474 YDRFPSPQRDDQDEIGALVSYLGPSTSPHRSRSPETWPLPHGGL-----523  
1071 YKGGGNPSK-----ALI-----KAPK---PMPTGSAIDKYKQEQR 1103  
524 -GRSVYAVEHL 533  
1104 SGRNIFAKFHV 1114

## RESULT 7

178879  
retinoblastoma binding protein 2 - human  
C:Species: *Homo sapiens* (man)  
C:Date: 17-Mar-2000 #sequence\_revision 17-Mar-2000 #text\_change 17-Mar-2000  
C:Accession: 178879; S16954  
R:Patasy, A.R.; Helin, K.; Dembski, M.S.; Dyson, N.; Harlow, E.; Vuocolo, G.A.; Hanobik  
Oncogene 8, 3149-3156, 1993  
A:Title: Characterization of the retinoblastoma binding proteins RBP1 and RBP2.  
A:Reference number: 158383; MUID:94020841; PMID:8414517  
A:Accession: 178879  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1722 <PAT>  
A:Cross-references: GB:S66431; NID:9435777; PIDN:AAB28544.1; PID:9435778  
R:Deleo-Jones, D.; Huang, P.S.; Jones, R.E.; Haskell, K.M.; Vuocolo, G.A.; Hanobik, M.G.  
Nature 352, 251-254, 1991  
A:Title: Cloning of cDNAs for cellular proteins that bind to the retinoblastoma gene pro  
tein.  
A:Reference number: S16953; MUID:91312450; PMID:1857421  
A:Accession: S16954  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1102-1562, 'KKK' <DEF>  
C:Genetics:  
A:Gene: GDB:RBP2  
A:Cross-references: GDB:119548; OMIM:180280  
A:Map position: 3q21-3qter  
C:Superfamily: human retinoblastoma binding protein 2

Query Match 4.4%; Score 136; DB 1; Length 1722;  
Best Local Similarity 19.1%; Pred. No. 2.1;  
Matches 146; Conservative 120; Mismatches 254; Indels 244; Gaps 38;  
Db 14 TTANFLQRHQSEORRELKIVETESTMDITIGSKQPOFLKSIDELAFAVAVETFK 73  
796 SVAQLLSK---KQKHQSP---DSGRTRTKLVSELK-AFVQQLFSLPCVISQARQVK 847  
74 RQFDLOKHIESIENAI-----DSKLESNGVVVAARNNNFHQMLSPRRNNVS- 122  
848 NLLDVEEFHERAQEAMMDTPDSSKLQ-----MLIDWGSSLYVELPELRLKQLQQAARW 903  
123 --ETVTVSQPSQIIVPETSNKPEGG-----RMCEL-----MCS 154  
904 LDEVRLTSDPQVTLDMVKLLDSGVGLAPHAVKAMAEQLLTVSERWEEKAKVCL 963  
155 KGLRKYIYANISQAKLMEEIPSAKLAKEPAKFVLDCTGKFLQGRRAFTK--ESPMS 212  
964 QARPHSVASLESIVNEAKNIPAPL-----PNVLSL-----KEALQKAREWTAKVEAIQSG 1014  
213 ARQVSLILLESFLMPDRGKKG---VKIESWIKDEASTAAV-AWRKRLMTREGGLAAAEKW 268  
1015 SNAYLEQLSEL-----SAKGRPIPVRLREALPOVESQVAAARAWRRT-----GRFLKKN 1065  
269 DARGLLLVACFGVPSNFRSTDLDIRMSGNEIAGALKRSQFLVPMVSGIVESSIKRG 328  
1066 SSHTLLQVLS-----PRTD---IGVVGS---GKNRRK---VKELIEKEKED 1104  
329 MHIALEMYVTFGMEDKFAALVLTSFLKMSKESFERAKRKAQSPFLAFKEAATKQLAVLS 388  
1105 LDLEPLSDLEE-GLEETRDYAMVAVFK-----BRQKEIEAMHSLRAANLAKMTMYD 1156  
389 SVMQ-----CMETHKLDPAKLPQWQKE-----412  
1157 RIEVEKFCIKRTASGFMFLQCELCQWFFNSCVPLPKSSSQKSGSQAKVEKFLCPLCM 1216  
413 -----QIVSLEKOTLDK-----EMEEKAR-----SLSLMEE 440  
1217 RSRPRLETILSLVSLQKLPVRLPSGEALQCLTERAMSQDRARQALATDELSSALAKL 1276  
441 AALAKRMYNQ-----QIKRPRLSP---MMPPVTSSYSPI-----473

Db 1277 SVLSQRMWQAAREKTEKIISAELOKAAANPDLQHLPSFOQSAPNRVSVSSSPRCQM 1336  
QY 474 -YRDRSFPQRD-----DDQDEISALVSSYLGPS-----TSFPHRSRSPYVPLPHG 521  
Db 1337 DYDDDETDSDIREYGYDMKDTASVKSSSLEPNLFCDBEIPKBEVVTHMTAPS- 1395  
QY 522 GLGRSVAYEHL---APNSYSP---GHGRLHRYQ---SPSLVHGQRHP-LOYSPPIHQ 571  
Db 1396 -----FCAEHAYSSASKSCSVFFGKSGSTPRKQPSPLVPSRLEPPVLELSPGAKA 1449  
QY 572 --QQLPYG-----IQRYVR-----HSPSEERYLGLSNQRS 599  
Db 1450 LSELMMVGLLEVSLEDTCHIRWILQATHPPSDFRFLHIMEDDS 1493  
  
RESULT 8  
A86188  
hypothetical protein [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C;Accession: A86188  
R;Theologos, A.; Ecker, J.R.; Palm, C.J.; Pederspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
C;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
C;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A;Reference number: A86141; MUID:21016719; PMID:11130712  
A;Accession: A86188  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-841 <STO>  
A;Cross-references: GB:AE005172; NID:G2388564; PIDN:AAB71445.1; GSPDB:GN00141  
C;Genetics:  
A;Map position: 1  
  
Query Match 4.3%; Score 134; DB 2; Length 841;  
Best Local Similarity 20.8%; Pred. No. 1;  
Matches 118; Conservative 74; Mismatches 196; Indels 180; Gaps 21;  
  
QY 20 LQHQSEQRRLPKIVETESTSMDITICQ-----SKQPLKSIDEL-----AAPS 66  
Db 132 LQK-QSASELEKLSIDERYKTDALLSQALSONSLVLEPKLSLEELSEKVSLEKLSALI 190  
QY 67 VAVETKQFDDLOKHLESTENAIKSLKESNGVLAARNNHPQM----- 112  
Db 191 VAEERGKSSIQMOYOEKV-----SKLESSLNQSSARNSLEEDRLALQGAHEDIG 245  
QY 113 -LGPPRN-----NVSVETTVTSQPSQEIYVETS 141  
Db 246 NVSTKSGVELQGLFQTSQKLEKAEKLDLEAIQVNSLSLEATLSVAMEKRDLSNLN 305  
QY 142 KPEGGRMELMCKSGRLKYIYANISQA-----KLMEIPALKLA 182  
Db 306 -----AVMEKLSSEERLEKQAREIDEATRSIEALHKKSELVKQKTMEDFSRDETEA 360  
QY 183 KEPAKFVLDICIGFYK-QGRRAFTKESPMSSARQVSLILLESFLLMPORGKGVKI---E 238  
Db 361 KSLTEKSKDLEEKIRVYEGKLAECAGSLSLQBELQSSAENELLADTNNQLKIQELE 420  
QY 239 SWIKDEAETAANVRKELMTEGGAAAEKMDARGILLLVACFGVPSNFRSTDLILLRMS 298  
Db 421 GYLDSEKET-----ATKEL-----NQKDTAKDLITKL 448  
QY 299 GSNEIAGALKESQPLVPMVSGIVESSIKRGHMTALEMVTYFGMEDKFGAALVLVSFLKM 358  
Db 449 KSHD--NVEIEHKQVLEASGVADT---RKVEVE-----EALLKNTLES 488  
QY 359 SKESFER-----AKRKAQSPILAFKAATQKLAVLSSVQNCMETHKLDPAKELFGWQIK 411

Db 489 TIEELEKENGDLAEVNIKNOKLANQSGETDDQAKLSV---LEAEKYQQAELQ----- 540  
QY 412 EQIVLEKEDTTLQDKEMEEKARSLSLMEEAALAKRMYNQOIKRPRJSPMPPVPTSSSYS 471  
Db 541 ---ITIEDLTQLTATSERLERLSQISSLEE---EKQVN-----BIYQSTKNELV 583  
QY 472 FYRDRSFPQSQRDDQDQ---DEISALVS 495  
Db 584 KLQAQLQVDKSKSDDMVMSQIEKLSALVA 611  
  
RESULT 9  
A26655  
myosin heavy chain [similarity] - slime mold (Dictyostellium discoideum)  
N;Contains: myosin A1Pase (EC 3.6.4.1)  
C;Species: Dictyostellium discoideum  
C;Date: 05-Oct-1988 #sequence\_revision 05-Oct-1988 #text\_change 19-Apr-2002  
C;Accession: A26655; A24728; S00250  
R;Warrick, H.M.; De Lozanne, A.; Leinwand, L.A.; Spudich, J.A.  
Proc. Natl. Acad. Sci. U.S.A. 83, 9433-9437, 1986  
A;Title: Conserved protein domains in a myosin heavy chain gene from Dictyostellium disc  
A;Reference number: A26655; MUID:87092266; PMID:3540939  
A;Accession: A26655  
A;Molecule type: DNA  
A;Residues: 1-2116 <WAR>  
A;Cross-references: GB:M14628; GB:M11938; NID:G167834; PIDN:AAA33227.1; PID:G167835  
R;DeLozanne, A.; Lewis, M.; Spudich, J.A.; Leinwand, L.A.  
Proc. Natl. Acad. Sci. U.S.A. 82, 6807-6810, 1985  
A;Reference number: A24728; MUID:86016788; PMID:3901008  
A;Accession: A24728  
A;Molecule type: mRNA  
A;Residues: 2035-2116 <DEL>  
R;Wagle, G.; Noegel, A.; Scheel, J.; Gerisch, G.  
FEBS Lett. 227, 71-75, 1988  
A;Title: Phosphorylation of threonine residues on cloned fragments of the Dictyostellium  
A;Reference number: S00250; MUID:89112226; PMID:2828113  
A;Accession: S00250  
A;Status: nucleic acid sequence not shown  
A;Molecule type: DNA  
A;Residues: 1734-1893 <WAG>  
C;Comment: The rod domain is highly periodic, containing a pattern of 7-residue repeats  
C;Superfamily: myosin heavy chain; myosin motor domain homology  
C;Keywords: actin binding; ATP; coiled coil; hydrolase; nucleotide binding; P-loop; pho  
F;1-818/Domain: globular head <HED>  
F;89-747/Domain: myosin motor domain homology <MOT>  
F;179-186/Region: nucleotide-binding motif A (P-loop)  
F;819-2116/Domain: alpha-helical rod <ROD>

Query Match 4.3%; Score 134; DB 2; Length 2116;  
Best Local Similarity 20.6%; Pred. No. 3.9; Indels 182; Gaps 27;  
Matches 123; Conservative 82; Mismatches 210

QY 16 ANPLQHQSEQRRLPK-IVETESTSMDITIGQSKQFQFLKS-----IDELAAFSVAV 69  
Db 817 ARPLLKRRNFEKEIKEREILELKSNTLDTSTQDKLEKSLKDTESNVLQRLQKAEK 876  
QY 70 EYFKRQFD-----DLQGHIESIENAIKSLKESNGVLAARN----- 105  
Db 877 ETLKAMYDSKDALEAKQKRELEIRVEDMESELDK-----KLLENLQNKRSVEEKVRD 930  
QY 106 -----NNFQPMLSPPRRNNVSVETTVT-VSQPSQEIYVETSNRPEG 145  
Db 931 LBEELQEQKLTNLEKLLKKKYEELKMKVNDQCSQTSIRLEKIKDELQKEVEELTES 990  
QY 146 GRMCELMCKSKGL-----RKVIYANISD-----QAKLMEIS----- 175  
Db 991 --FSEESKDKGVLEKTRVRLQSELDLTVRLDSEFTKDSKESLLRQKKLEELKQVQEAAL 1048  
QY 176 -PSALKLAKPAKFLVLDICIGFYKQGRRAFTKESPMSSARQVSLILLESFLL-----MPD 229  
Db 1049 AETRAKLAQSAANKLQ--GE-YTELNEKFNSFVARSNVEKSKTLESQLVAVNNELDE 1105

QY 230 RKGKVKIESWIKODEATAVAMRK-RLMTGG-----LAAAEKWDAGLLLLVACFG 281  
DB 1106 EKORDALEK-KKALDAMLEENKQDLESTGGKSLYDLVKQBSDMALANQI-----1159  
QY 282 VPSNFRST-DLLDLIRSGSNEIAGALKRSQFLVPMVSGIVBESSIKRGHIEALEMVTYF 340  
DB 1150 --SELQSTIAKLEKIKSTLEGEVA-----RUGLEBAEQLAKSNVEKQKKVEL 1206  
QY 341 GMEKFSALVLTGFLKMSKESFERAKQAQSPILA-----FKEAATK-----382  
DB 1207 DLEDK-SAQLAEBT---AAQALDKLKKLEQELSEVOTQLSEANNKNVNSDSTNKHLET 1262  
QY 393 -----QIAVLSSVMQCMETHKLDPAKLPQWIKQEI-----VSLEKOT 421  
DB 1263 SFNNKLEAEQAKQALEKKRIGLSELEK-HVNEQLEEEKKQKESNKRKVDLEKEV 1320  
QY 422 LQDKEMEKEARSLSLMBEALAKRMVNNQIKRPLSPMPMPVTSYSSPIYDRS 478  
DB 1321 SELKQDIEEVASKAVTEAKNKKSELEIKR-----QYADVSS-----RDKS 1365  
RESULT 10  
T13564  
Microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)  
N:Alternate names: hypothetical protein EG:49E4.1  
C:Species: Drosophila melanogaster  
C>Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
A:Accession: T13564  
R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.  
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.  
A:Reference number: Z17689  
A:Accession: T13564  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-5327 <SPA>  
A:Cross-references: EMBL:AL031128; PIDN:CAA20006.1  
C:Genetics:  
A:Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1  
A:Note: EG:49E4.1  
C:Superfamily: Drosophila 576K microtubule-associated protein homolog

Query Match 4.3%; Score 132; DB 2; Length 5327;  
Best Local Similarity 20.7%; Pred. No. 20;  
Matches 119; Conservative 103; Mismatches 219; Indels 134; Gaps 27;  
QY 7 TVAAOPTTA-NPLLRHQS-----EQRRRLPKIVETESTMDITIGQSQPOFLKSID 60  
DB 3976 TVTKQSETTLFETLTSKVESKVEVLESSVKQVEKVTQSVKQATTTVDLSQLTKSSZ 4035  
QY 61 ELA-----AFSAVETFRQFDD-----LQKHESI-----ENADS 92  
DB 4036 QLTEIKSVLDTNISVNTLNFSTAVETIEKKVQDVTEKVIKATEHVSEHVTTTGESSTET 4095  
QY 93 KLESNGVLA--ARNNHQHPMLSPRNVSVEVTTVTSQSPQIVPETSNNKPEGGRV-- 148  
DB 4096 SOEKSLLDGTSELRETHITTVGSP-----EFTVICERDEPVLADIKEEDHEHRFSP 4149  
QY 149 -----CELMCSKGLRKTYIANISQAKLMBEIPFSAKLAKPAKFLVDCIGKF-----196  
DB 4150 PSDVDKAAIIPQWNP-LSPREEVAKIVADVAKVLKSDK-----ITDIIPDFDERQLE 4204  
QY 197 -VLOGRAFTKESPMSSARQVSLI-----LESFLAMPDRGKGVKIESWIKODEATAV 250  
DB 4205 EKLKSTADEESKTRDEKSLSVKVEIESKSPDQSGPISIEE--KQIKES-- 4260  
QY 251 AWRKRLMTGEGGLAAAEKMDARGLLLVACFGVPSNFRSTDLDDLRMSGNEIAGALKRS 310  
DB 4261 --EKQLRQGLASSRPESVASQPESVPS---PQSAASHHEKVEVSESHK-AEKSRRP 4314  
QY 311 QFLVPMV-----SGVIESIKRG-----MH-TEALEMVTYFGMEDKPSAAL 350

DB 4315 ESVAQSVSEKMDKTSRPASTSQFSTKGEDETTBSLLHSLTTTETVETKQMEKSSPES 4374  
QY 351 VLTSFLKMSKESFERAKQAQSPILAKPAKAAATQQLAVLSSVMQCMETHKLDPAKLPQWQI 410  
DB 4375 VTSVTKSTVLS-----SQSTVQLREESTSE--SLSSSLKVEDSSRRSLSL-----L 4421  
QY 411 KEQ-----IVSLEKDTLQLDKEME-----EXARSLSLMEE--AALAKMYNQOIKRPL 457  
DB 4422 AEKGGIATNTSLKEDTSASASQLEELLVQSECSSESIVSEICTSTQAKS-NKEIKDARE 4480  
QY 458 SPMEPPVTSYSSPIYDRSFPSPDDDDQDEISA 492  
DB 4481 TKVTSQFTTTS-----SATKDDSLKETVA 4505  
RESULT 11  
S49461  
Synaptonemal complex protein 1 - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 05-Nov-1999  
A:Accession: S49461; S59599  
R:Julien, S.; Luc, M.; Francois, C.  
A:Submitted to the EMBL Data Library, October 1994  
A:Description: Cloning and sequencing of the murine SCPI cDNA.  
A:Reference number: S49461  
A:Accession: S49461  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-993 <JUL>  
A:Cross-references: EMBL:Z38118; NID:G1360015; PIDN:CAA86262.1; PID:G558603  
R:Sage, J.; Martin, L.; Curin, F.; Rassoulzadegan, M.  
Biochim. Biophys. Acta 1263, 258-260, 1995  
A:Title: cDNA sequence of the murine synaptonemal complex protein 1 (SCPI).  
A:Reference number: S59599; MUID:96004899; PMID:7548215  
A:Accession: S59599  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-993 <SAG>  
A:Cross-references: EMBL:Z38118; NID:G1360015; PIDN:CAA86262.1; PID:G558603  
Query Match 4.2%; Score 131; DB 2; Length 993;  
Best Local Similarity 19.8%; Pred. No. 2;  
Matches 122; Conservative 97; Mismatches 222; Indels 176; Gaps 29;  
QY 13 TTTANPLQRHQBQRRLPKIVETESTMDITIGQSQPOFLKSIDELAAP-----SVA 68  
DB 381 TCTLEELL---RTEQQRLE-----KNEDQLKLITVBLQKKS---NELEMTKFNKNKEVE 429  
QY 69 VETPK-----ROPDDLQKHIESIENADSKLESNGVLAARNNNPHQPMSPRNNSV 122  
DB 430 LEEKNTLAEDQKLLDEKKQVEKLABELQKEQELTFLLETRKEVH-----DLQE 480  
QY 123 ETTVT-----VSQPSQEIIVPETSNNKPEGGRVCELMCSKGLKTYIANISQAKLM 172  
DB 481 QVTVTKTSEQHYLKQVEEMKTELEKEKLNTELTAASCDML-----LLENKKEV 528  
QY 173 EEIISALKLAKPAKFLVDCIGKFYQGRFAFTKESPMSSARQVSLIILESFLMPDRGK 232  
DB 529 QEASDMALEKHQEDLIINC-----KKQEERLLKQIENL-----562  
QY 233 GKVIIESWIKODEATAVAVAKRLMTGEG-----LAAAEKMDARGL-----LL 275  
DB 563 --EEKEMHLRDELESV---RKPFIOQGDVCKLDKSEB-NARSTICEVLKKEKQMKIL 615  
QY 276 LVACFGVPSNF-RSTDLDDI-----IRMSGNEIAGALKRSQFLVPMVSGIVSISIKR 327  
DB 616 ESKNNKQVENSCKRIEHLHENTKTKKSSAEIK-QUNAVEIKVSKLELESTKGR 674  
QY 328 GMHTEALEMVTYFGMEDK-----FSAALVLTSLFKMSKESFERAKQAQSPILAF 376  
DB 675 ---FEEMTNVQKIEENKTSSEKGLGEVEKAKATVDEAVKLQKEIDLRCQHIAEWAL 731  
QY 377 KEAATKQ-----LAVLSSVMQCMETHKLDPAKLPQWIKQEIQVLSLEKDTLQLD 425

Db 732 MEKHKQYDKIVEERDSGLYKREQSSAKIALETLS--NIRNELVSLKK--QLE 786  
Qy 426 KEMEEKARSLSLWEAALAKMYNQIKRPLSPWMPPTSSSVSYPIYDRGPPSQ-- 483  
Db 787 IEKEKEKLMKAKENTALK---DKDKKIQASLESPEATSWKFD---SKTTPSQNIS 839  
Qy 484 -----DDQDEISALVSSYLGSTSPPHRSRSPYIMVPLPHGGGLGRSVAYEH 532  
Db 840 RLSSMSDGSXDNRLNRASAKSIL-PTTV-----TKETVTKTP---TKSIYQRE- 887  
Qy 533 LAPNSYSPGHGRLHRQ 549  
Db 888 ---NKYIPTGGSNKKRK 901  
RESULT 12  
SCPI protein - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 08-Oct-1999  
C:Accession: S28061  
R:Neuwissen, R.L.J.; Offenberg, H.H.; Dietrich, A.J.J.; Rieswijk, A.; van Iersel, M.; HEMBO J. 11, 5091-5100, 1992  
A:Title: A coiled-coil related protein specific for synapsed regions of meiotic prophase  
A:Reference number: S28061; MUID:93099884; PMID:1464329  
A:Accession: S28061  
A:Molecule type: mRNA  
A:Residues: 1-946 <NEU>  
A:Cross-references: EMBL:X67805; NID:G57212; PIDN:CAA48006.1; PID:G57213  
C:Genetics:  
A:Gene: SCPI  
C:Keywords: DNA binding  
Query Match 4.2%; Score 129; DB 2; Length 946;  
Best Local Similarity 19.1%; Pred. No. 2.5; Mismatches 235; Indels 146; Gaps 25;  
Matches 115; Conservative 106;  
Db 13 TTTANPLLQHQHSEQRRLPKIVETESTSMDITIGQSKQFQFLKSIDELAAF----SVA 68  
Db 334 TCTLELL--RTEQRL--NNEOLKLTWELQKS---SELEMTKFKNNKEVE 382  
Qy 69 VETFK-----QFDDLOKHIESIENADSKLESNGVLAARNNHPOMLSPNNVSV 122  
Db 383 LEBLKTLAEDOKLLDEKKQVEKLAELQGEQELTFLLOTREKH-----DL 431  
Qy 123 ETTVTVSQSQE-----IVPETSNGKPGGRM--CELMCSKGLRKYIYANISDAKLMBEIP 176  
Db 432 EVQVTVTKTSEHYLKQVEEMTELEKEKLNIELTANSDMI-----LLENKLVQEAS 485  
Qy 177 SALKAKEPAKFVLDICIGFYLGRRAPTKESPMSARQVSLILLESFLMPDRGKGVK 236  
Db 486 DMVLELKHQEDIIINCKQBEERMLQIETLEBKEMNLRDELESVRKFTIQGDEVKCKL- 544  
Qy 237 IESWIWDEAETAAVARKLMTTEGGLAAREKVDAGLLLVACFQVSNF--RSTDLDDL 294  
Db 545 -----DKSEENARSIEYEV-----KKEOKMILEKNCKKQIENKKNIEEL 589  
Qy 295 -----IRMSGSN-----ETAGALKRSQFLVPMVSGIVESSIKRGHMI 331  
Db 590 HQENKALKKSSAENKQLNAYEIKVKNLELASTKQKFEEMI-----NNYQKTEI 641  
Qy 332 EALEMVTGMDKFAALVLTSFLKMSKESFERAKRAQSLAPKEAATVQ----- 383  
Db 642 KKTISEKLLGEVEK--AKATVDEAVKLEQIEDLRQCHTAENVALMEKHQYDKIVEER 699  
Qy 384 ---LAVLSSVMQCMETHKLDPAKELPGWQIQEIVSLEKDTLQLDKEMEERKARSISLME 440  
Db 700 DSLEGLYKREQSSAKIALETLS--NIRNELVSLKK--QLEVEKEEKLKMEQEN 754  
Qy 441 AALAKMYNQIKRPLSPWMPPTSSSVSYPIYDRGPPSQ-----DDQ 487  
Db 755 TAI---LTDKKDKKIQASLESPEATSWKFD---SKTTPSQNISRLSSMSDGSKKNR 807

Qy 488 DEISALVSSYLGSTSPPHRSRSPYIMVPLPHGGGLGRSVAYEHAPNSYSPGHGRLH 547  
Db 808 DSLRASAKSILSTTV-----TKETVTKTP---TKSIYQRE-----NKYLTGGSNKK 852  
Qy 548 RQ 549  
Db 853 RK 854  
RESULT 13  
B48666  
cell proliferation antigen Ki-67, short form - human  
C:Species: Homo sapiens (man)  
C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 07-May-1999  
C:Accession: B48666  
R:Schlueter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Gerdt  
J. Cell Biol. 123, 513-522, 1993  
A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, ubi-  
ins.  
A:Reference number: A48666; MUID:94043435; PMID:8227122  
A:Accession: B48666  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-2897 <SCH>  
A:Cross-references: EMBL:X65551  
C:Superfamily: kinase interaction domain homology  
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat  
F:29-91/Domain: Kinase interaction domain homology <KIH>  
Query Match 4.1%; Score 127.5; DB 2; Length 2897;  
Best Local Similarity 20.0%; Pred. No. 16;  
Matches 137; Conservative 92; Mismatches 228; Indels 229; Gaps 32;  
Qy 11 OPTTANPLLQHQHSEQR--RELPKIVETESTSMDI-----TIGQSKQFQFLK 57  
Db 505 ESKTVSTVNSGRSTEFNRNQKLPVSKSEETNTEIVECILKRGOKATILQORREGEMK 564  
Qy 58 SDELAFAFSAVETPKRQFDDLOKHIESIENADSKLESNGVLAARNNHPQ---PMLS 114  
Db 565 EIE-----RPFETYKENIELKEN--DEKMA-----MKRSRTWGQKCAPM-- 602  
Qy 115 PRNNVSVETTVTSQSQEIVPETS-----KPEGGRMCELMCSKGLRKY 160  
Db 603 -----SDLTDLKSLPDTLMKDTARGONLLQTDHAKAPKSEKIKTKNPOQSLQPEP 655  
Qy 161 IVANISDAKLMEEIPSAKLAKPAKFPVLDICIGF-YLQGRRAFTKESPMSSARQVSL 219  
Db 656 INTPHTTKQOL-----KASLGKGVKKEELLAAGKFTTSGETTHTHREPAGDKSI-- 706  
Qy 220 ILESFLMPD-----RGKGVKIESWIKDEAATA--VAWRKELMTEGGLAAREKMDAR 271  
Db 707 --RTPKESPQIILDPAARVTGKWPPTKPEEAOSLEDLAGFKELFQTPGFPSEESMTDEK 764  
Qy 272 GLLLVACFGVP-----SNFRSTDL-----LDLIRMSGSNBIAGALKRSQF 312  
Db 765 ---TTKIACKSPPEVSDPTSTKQMPKESLEKADVEEFLALRLKLTPS---AG---KAM 815  
Qy 313 LVPVMSGIVESSIK--RGWHIEALEMVTTFQMDKFAALVLTSFLKMSKESFERAKKA 370  
Db 816 LTPKPAAGGDEKDIKAFMGTPVQKLDLAGT-----LFGSKRLQTPKPKA 859  
Qy 371 OS--PLA-PKE-----AATKQ----- 383  
Db 860 QALEDLAGFKELFQTPGHTHELVAGKTKTKPCDPSQSDPVDTPSTTKQPKRSIRKADV 919  
Qy 384 ----LA-----VLSSVMQCMET-----HKLDPKELPGWQIQEIVSL 417  
Db 920 EGELACRLNLMPSAGKAMHTPKPSVGBEKDIIIFVGTVPVQKLDLTENLTGSKRPQTPKE 979  
Qy 418 EKDITQLDKEMEERKARSISLMEBAALAKMYNQIKRPLSPWMPPTSSSVSYPIYDR 477  
Db 980 EAALEDITGFKELFQTPGHTHEEVAAG-----KTKMPCSSPPESADTPTSTRQP 1032

QY 478 SPSQDDDDQDEISAL--VSSYLGPSTSPFRRSRSPRYMVLPHGGLGRSVAYEHLAP 535  
Db 1033 KTFLEKRDVQKELSAKLTQTSGETT---HTDK-----VP---GGEDKSINAFRETAK 1080  
QY 536 NSYSPGHGRLHRLHROYSPLVHGQRHP 561  
Db 1081 QKLDP-----AASVTGSKRHP 1096  
RESULT 14  
A48666  
cell proliferation antigen Ki-67, long form - human  
C:Species: Homo sapiens (man)  
C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 08-Oct-1999  
C:Accession: A48666  
R:Schueter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Gerde  
J. Cell Biol. 123, 513-522, 1993  
A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiqui-  
tous protein  
A:Reference number: A48666; MUID:94043435; PMID:8227122  
A:Accession: A48666  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-3256 <SCH>  
A:Cross-references: EMBL:X65550; NID:G415818; PIDN:CAA46519.1; PID:G415819  
C:Superfamily: Kinase interaction domain homology  
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat  
F;29-91/Domain: Kinase interaction domain homology <KIH>  
Query Match 4.1%; Score 127.5; DB 2; Length 3256;  
Best Local Similarity 20.0%; Pred. No. 19; Mismatches 228; Indels 229; Gaps 32;  
Matches 137; Conservative 92; Nucleotide 228; Indels 229; Gaps 32;  
QY 11 QPTTANPLLRHQSRQR--RELKIVETESTMDI-----TIQSKQPOFLX 57  
Db 864 EFSKTVSTVNRSGSRTEFNQIKLPESKEETNTEIVEICLKGRQKATLLQORREGEMK 923  
QY 58 SIDELAAPSVAVETFKQFDDLOKHIESIENAIKLESNGVLAARNNHQ---PMLS 114  
Db 924 EIB-----RPPETKEMIELKEN---DEKKA-----WGRSRTWGQKCAPM-- 961  
QY 115 PPRNNVSVETTVTSOPSOEIVPETSNT-----KPEGGRMCELMSCKGRKY 160  
Db 962 -----SDLTDLKSLPDELAKDITARGQNLQATQDHAKAPKSEKIKTKMPCQSLQPEP 1014  
QY 161 IYANISQAKLMEEIPSAKLAKPAKFVLDICIGF--VQGRRAFTKESPMSSARQVSL 219  
Db 1015 INTFTTKQOL-----KASLGKVGKESLLAVGKFTRTSGTTHREPAAGDKSI--- 1065  
QY 220 ILESFLLMPD-----RGKGVKIESMTKDAETAA--VAVKRLMTGEGGLAAAEKMDAR 271  
Db 1066 --RTFKESPKQILDPAARVTGMKNFRTPEBAQSLDLAGFKELFQTPGSEESMTDEK 1123  
QY 272 GLLLLVACFGVP-----SNFRSTDLL--LDLRMSGSNEIAGALKRSQF 312  
Db 1124 --TTKIACKSPPPESVDTPTSTKQWPKRSLRADVEEBFLALRLKLTGS---AG----KAM 1174  
QY 313 LVPWVSGIVESSIK--RGWHIEALEMVTTFMGEDKFSALVLTSLFKMSKESFERAKKA 370  
Db 1175 LTPKAGGDEKDIKAPMGTPVQKDLAGT-----LPGSKQLQTPKEKA 1218  
QY 371 QS--PLA-FKE-----AATKQ----- 383  
Db 1219 QALEDLAGFKELFQTPGHTTEELVAAGTKTKIPCDSPQDPVDTPTSTKQPKRSIRKADV 1278  
QY 384 ----LA---VLSSVMCMET-----HKLDPAKELPGWQKEQIVSL 417  
Db 1279 EGGLEACRNLMPAGAMHTPPSVEEEDKIIIFVGTVPQKDLDTENLTGSKRPQTPKE 1338  
QY 418 EKDTQLDKMEMEEKARSLMEBAALAKRMVNNQIKRPRLSMEMPPVTSSSYPIYDR 477  
Db 1339 EAQALDGLTGFELFQTPGHTTEEVAAG-----KTKMPCSSPPESADTPTSTRQP 1391

QY 478 SPSQDDDDQDEISAL--VSSYLGPSTSPFRRSRSPRYMVLPHGGLGRSVAYEHLAP 535  
Db 1392 KTFLEKRDVQKELSAKLTQTSGETT---HTDK-----VP---GGEDKSINAFRETAK 1439  
QY 536 NSYSPGHGRLHRLHROYSPLVHGQRHP 561  
Db 1440 QKLDP-----AASVTGSKRHP 1455  
RESULT 15  
T19898  
aryl hydrocarbon receptor AHR-1 homolog - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jul-2000  
C:Accession: T19898; T42398  
R:Steward, C.  
submitted to the EMBL Data Library, October 1996  
A:Reference number: Z19192  
A:Accession: T19898  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-602 <WIL>  
A:Cross-references: EMBL:Z81048; PIDN:CAB51463.1; GSPDB:GN00019; CBSP:C41G7.5  
A:Experimental source: Clons C41G7  
R:Fowell-Cotman, J.A.; Bradfield, C.A.; Wood, W.B.  
Proc. Natl. Acad. Sci. U.S.A. 95, 2844-2849, 1998  
A:Title: Caenorhabditis elegans orthologs of the aryl hydrocarbon receptor and its heter-  
odimeric ligand  
A:Reference number: Z22156; MUID:98169459; PMID:9501178  
A:Accession: T42398  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-602 <POM>  
A:Cross-references: EMBL:AF039570; NID:G2828113; PIDN:AAC00000.1; PID:G2828114  
C:Genetics:  
A:Gene: ahr-1; C41G7.5  
A:Map position: 1  
A:Introns: 13/2; 33/2; 76/1; 210/1; 234/3; 270/2; 303/2; 338/1; 386/2; 436/2; 541/1; 557  
Query Match 4.1%; Score 126.5; DB 2; Length 602;  
Best Local Similarity 19.0%; Pred. No. 1.9; Mismatches 243; Indels 171; Gaps 29;  
Matches 120; Conservative 99; Nucleotide 243; Indels 171; Gaps 29;  
QY 5 PPTVAAQPTTANPLLRHQSRQRRELKIVETESTMDITIQSKQPOFLKSIDELAA 64  
Db 17 PP-----KQLTWNTN-----SKRHERLING--ELETVAMLLPYDSST----ISRLDKLSV 60  
QY 65 FSVAVETFKRQFDDLOKHIESIENAIKLESNGVLAARNNH--HOPMLSPFRNVSV 122  
Db 61 LRLAY-----SFLQCKAHFOACLH--NSQFLSAGFPMTSHSYSQPHPTI--PFSNKVPT 111  
QY 123 ETTVTVSOPSOEIVPETSNT-----KPEGGRMCEL-----NCSKGLRKYI----- 161  
Db 112 IPDLRIGTPMLD--PEESNFEISLKGFLVLDNDNGEIIYASENVENYLGHQSDVL 169  
QY 162 -----YANISQAKLMEEIPSAKLAKPAKFVLDICIGKFIQGRRAFTKESPMASARQ 215  
Db 170 HQPVYDLIHSERDDIRQQLDSNFHPTTSSASNOFD-----VFAPQNSKYLERN 218  
QY 216 VSLILLESFLAMPGRGKGVKIESWIKDEAETAFAVAVKRLMTGEGGLAAAEK--DARGL 273  
Db 219 VN-----AFRCILLDNTCGFLID-----MRGLKMSLHGLPSSVYVWGRATSGP 261  
QY 274 LLLVACFGVP-----SNFRSTDLLDIRMSGSNEIAGALKRSQFLVPMVSGIVESSIKR 327  
Db 262 VLGMICVCTPFPVPESTSLASEDMI----LTKHQLDGALVSDQKYYEMLEIDETDLP 317  
QY 328 GMHIEALEMVTTFMGEDKFSALVLTSLFKMSKESFERAKKQSPALAKFAEATKQLAVL 387  
Db 318 PLI-----NLVH---VEDA-----VCVAEAKKAIKNGSGLLVYRLVSTKTRTY 360  
QY 388 SSVVMCMETHKLDPAKELPGWQKEQIVSLEKDTQLQDKMEMEEKARSLMEBAALAKRM 447

361	FVQSCRFYKNSKPESI---	GLTHRLNVEGTMLLEKSTLIKALLSFDDSF-----	411	
Db				
448	YNQIKRPLSPMEPPVTSSSYPIYDRSFPSQRDDQDEISALVSYLGFTSPF--	505		
Qy				
412	-----LQSER-----	NLQSTAAFLP-----	PVLKDDQDCLEPSTNSLFPSPVPVTP	454
Db				
506	-----HRSRRSPYMWPLPHGGLGRSVAYAYEHLAPNSYSPGHGHLRHQYSPSLVQ	558		
Qy				
455	TTTKANRRRKENSHEIVTIP-----	SIPIPTHFDMQMFDPSPWNHGVH--	PPAMPHDV	505
Db				
559	RHPLQYSPPIHGQQQLPYGIGRVVYVHSPSEERY	591		
Qy				
506	YHLYTYP-----	TPHPPTGTWGY	524	
Db				

Search completed: August 14, 2003, 10:40:50  
Job time : 46 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 14, 2003, 10:28:09 ; Search time 25 seconds  
(without alignments)  
1145.571 Million cell updates/sec

Title: US-09-890-475-1

Perfect score: 3104

Sequence: 1 MSNYPPTVAQPTTTANPLL.....RYLGLSNQSPRSNSLDPK 609

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3104	100.0	609	1	FRI_ARATH
2	144	4.6	724	1	Q9fdw0 arabidopsis
3	136	4.4	1722	1	P29375 homo sapien
4	134	4.3	2116	1	P08799 dictyosteli
5	131	4.2	993	1	Q62209 mus musculus
6	129	4.2	987	1	Q03410 rattus norv
7	127.5	4.1	3286	1	P46013 homo sapien
8	126.5	4.1	1920	1	P48725 mus musculus
9	126	4.1	694	1	P91240 caenorhabdi
10	126	4.1	2230	1	Q13439 homo sapien
11	125.5	4.0	6895	1	Q8wxh0 homo sapien
12	125	4.0	1128	1	P32873 saccharomyc
13	124.5	4.0	1197	1	Q92m7 homo sapien
14	123.5	4.0	5120	1	Q9p36 gallus gall
15	123	4.0	485	1	Q9bxx5 homo sapien
16	122.5	3.9	1336	1	Q9C0J8 homo sapien
17	122.5	3.9	2212	1	Q93074 homo sapien
18	122	3.9	962	1	Q9Xhr2 zea mays (m
19	122	3.9	1325	1	P55937 mus musculus
20	120.5	3.9	1875	1	Q03455 saccharomyc
21	118.5	3.8	2044	1	P91620 drosophila
22	118.5	3.8	2064	1	P91621 drosophila
23	117.5	3.8	1060	1	P28025 xenopus lae
24	117.5	3.8	2842	1	Q70478 rattus norv
25	116	3.7	1087	1	P94212 homo sapien
26	115	3.7	1499	1	Q13017 homo sapien
27	115	3.7	1690	1	Q9v1e5 drosophila
28	114.5	3.7	782	1	Q10030 caenorhabdi
29	114.5	3.7	943	1	P34408 caenorhabdi
30	114.5	3.7	1940	1	P12847 rattus norv
31	114.5	3.7	1972	1	P35749 homo sapien
32	114.5	3.7	2404	1	Q9Qx47 mus musculus
33	114.5	3.7	4473	1	Q91i55 cricetus

34	114	3.7	1125	1	YB62 SCHPO
35	114	3.7	1433	1	Y310 HUMAN
36	114	3.7	1972	1	MYHB_MOUSE
37	113.5	3.7	1025	1	ADAL_YEAST
38	113.5	3.7	1505	1	CUT2_HUMAN
39	113.5	3.7	1861	1	MAP2_RAT
40	113.5	3.7	1940	1	MYH3_CHICK
41	113.5	3.7	1940	1	MYH3_HUMAN
42	113.5	3.7	2349	1	TPR_HUMAN
43	113	3.6	554	1	RCN1_RICPR
44	113	3.6	1790	1	USO1_YEAST
45	113	3.6	1938	1	MYH4_RABIT

#### ALIGNMENTS

#### RESULT 1

FRI\_ARATH STANDARD; PRT; 609 AA.  
 ID Q9FDW0; O65274; AC  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE FRIGIDA protein.  
 GN FRI OR Ar4G00650 OR F6N23.25.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;  
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND VARIANTS EDI-0; SF-2 AND SHAKHDARA.  
 RC STRAIN=cv. H51;  
 RX MEDLINE=20485641; PubMed=11030654;  
 RA Johanson U., West J., Lister C., Michaels S., Amasino R.M., Dean C.;  
 RT "Molecular analysis of FRIGIDA, a major determinant of natural  
 variation in Arabidopsis flowering time."; Science 290:344-347(2000).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=cv. Columbia;  
 RX MEDLINE=20083488; PubMed=10617198;  
 RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,  
 Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terry N.,  
 Harris B., Anserge W., Brandt P., Grivell L., Rieger M.,  
 Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,  
 Kreis M., Delsen M., Puigdomenech P., Watson M., Schmidheini T.,  
 Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,  
 Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,  
 Langham S.-A., McCullagh B., Bilham L., Robben J.,  
 Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,  
 Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,  
 Weitzenecker T., Bothe G., Ransperger U., Hilbert H., Braun M.,  
 Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,  
 Moeliman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,  
 Bernsner S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,  
 De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,  
 Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,  
 Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,  
 Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,  
 Borkova D., Bloeker H., Schaefer M., Grimm M., Loehner T.-H.,  
 Dose S., de Haan M., Maare A., Schaefer M., Mueller-Auer S.,  
 Gabel C., Fuchs F., Fortmann B., Grandrath K., Dauner D., Herzl A.,  
 Neumann S., Argirou A., Vitale D., Liguori R., Piravandi E.,  
 Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R.,  
 Schnabl S., Hiller R., Schmidt W., Lechary A., Aubourg S.,  
 Chedfor F., Cooke R., Berger C., Monfort A., Casacuberta E.,  
 Gabbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,  
 Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,  
 Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,  
 Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,  
 Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,

O14248	schizosacch
O15027	homo sapien
O08638	mus musculus
P38068	saccharomyc
Q14529	homo sapien
P15146	rattus norv
P02565	gallus gall
P11055	homo sapien
P12270	homo sapien
Q9zdy2	rickettsia
P25386	saccharomyc
Q28641	oryctolegus

RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,  
RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,  
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,  
RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,  
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,  
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,  
RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,  
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,  
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,  
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,  
RA Swaby J.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,  
RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,  
RA Chen E., Marra M., Martienssen R., McCombie W.R.,  
RT "sequence and analysis of chromosome 4 of the plant Arabidopsis  
thaliana";  
RL Nature 402:769-777(1999).  
CC -!- FUNCTION: Required for the regulation of flowering time in the  
late-flowering phenotype. Increases RNA levels of 'FLOWERING LOCUS  
C'. Variants with an early-flowering phenotype (including cv.  
Columbia, cv. Landsberg Erecta and cv. Wassilewskija) show loss-  
of-function mutations of FRI.  
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -!- CAUTION: The protein sequence is incorrectly predicted in the  
genomic sequence of cv. Columbia, but as the gene is defective in  
this cultivar it cannot be predicted at protein level correctly.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
-----  
DR EMBL; AF228500; AAC23415.1; -;  
DR EMBL; AF228499; AAC23414.1; -;  
DR EMBL; AF058919; AAC13626.1; ALT SEQ.  
DR EMBL; AL161472; CAB80874.1; ALT SEQ.  
KW Flowering; Coiled coil; Developmental protein; Nuclear protein.  
FT DOMAIN 409 409 COILED COIL (POTENTIAL).  
FT DOMAIN 55 55 COILED COIL (POTENTIAL).  
FT VARIANT 79 79 F -> I (IN CV. SHAKHARA).  
FT VARIANT 146 146 L -> I (IN CV. SF-2).  
FT VARIANT 148 148 G -> E (IN CV. EDI-O).  
FT VARIANT 148 148 M -> I (IN CV. EDI-O).  
SQ SEQUENCE 609 AA; 68443 MW; FE8572DF3E6726F CRC64;  
Query Match 100.0%; Score 3104; DB 1; Length 609;  
Best Local Similarity 100.0%; Pred. No. 6.5e-182;  
Matches 609; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MSNYPPTVAQPTTANPLLRHQSEQRRELPKIVETESTSMITIGSQKQPFLKSID 60  
DB 1 MSNYPPTVAQPTTANPLLRHQSEQRRELPKIVETESTSMITIGSQKQPFLKSID 60  
QY 61 ELAAPSVAVETKRPQDDLOKHESIEINAIDSKLESNGVLAARNPHQPLSPRRNV 120  
DB 61 ELAAPSVAVETKRPQDDLOKHESIEINAIDSKLESNGVLAARNPHQPLSPRRNV 120  
QY 121 SVETTVTSQPSQEIIVPETSINPEGRMCELMCSKGLKRYANIISDQAKLMEIIPSAK 180  
DB 121 SVETTVTSQPSQEIIVPETSINPEGRMCELMCSKGLKRYANIISDQAKLMEIIPSAK 180  
QY 181 LAKEPAKFLVDCIGFYLGRAFTKESPMSSARQVSLILLESFLLMPDRGKGVKIESW 240  
DB 181 LAKEPAKFLVDCIGFYLGRAFTKESPMSSARQVSLILLESFLLMPDRGKGVKIESW 240  
QY 241 IKDEATAAVANKRLMTGGGLAAEKMDARGLLLVACFGVPSNFRSTDLDLIRMSG 300  
DB 241 IKDEATAAVANKRLMTGGGLAAEKMDARGLLLVACFGVPSNFRSTDLDLIRMSG 300  
QY 301 NETAGALKRSQFLVPMVSGIVSESSIKRGHIEALEMWTYFGMEDKFSALVLTSLFKMSK 360

DB 301 NETAGALKRSQFLVPMVSGIVSESSIKRGHIEALEMWTYFGMEDKFSALVLTSLFKMSK 360  
QY 361 ESFERAKRKQSPFLAFKEAATKQALVLSVMQCMETHKLDPAKELFGWQIKQIVSLEKD 420  
DB 361 ESFERAKRKQSPFLAFKEAATKQALVLSVMQCMETHKLDPAKELFGWQIKQIVSLEKD 420  
QY 421 TLQLDKEMEKEARSLSLMEEAALAKRMYNQOIIRPRLSPMEMPPTVSSSYSPYDRSP 480  
DB 421 TLQLDKEMEKEARSLSLMEEAALAKRMYNQOIIRPRLSPMEMPPTVSSSYSPYDRSP 480  
QY 481 SORDDDODEISALVSSYLGSTSPFPHRSRSPYMWPLPHGGGLGRSVYAYEHLAPNSYP 540  
DB 481 SORDDDODEISALVSSYLGSTSPFPHRSRSPYMWPLPHGGGLGRSVYAYEHLAPNSYP 540  
QY 541 GGHRLHRQYSPSLVHGQRHLPQYSPPIHQOQQLPGVIGQVYRHSSEERYGLSNORSP 600  
DB 541 GGHRLHRQYSPSLVHGQRHLPQYSPPIHQOQQLPGVIGQVYRHSSEERYGLSNORSP 600  
QY 601 RSNSSLDPK 609  
DB 601 RSNSSLDPK 609  
RESULT 2  
ID HMNR HUMAN STANDARD; PRT; 724 AA.  
AC 075330; Q92767;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hyaluronan mediated motility receptor (intracellular hyaluronan acid  
binding protein) (Receptor for hyaluronan-mediated motility) (CD168  
antigen).  
GN HMNR OR IHABP OR RHAMM.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND CHARACTERIZATION.  
RC TISSUE=Breast carcinoma;  
RX MEDLINE=98264864; PubMed=9601098;  
RA Assmann V., Marshall J.F., Fieber C., Hofmann M., Hart I.R.,  
RT "The human hyaluronan receptor RHAMM is expressed as an intracellular  
protein in breast cancer cells."  
RL J. Cell Sci. 111:1685-1694(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Breast.  
RX MEDLINE=97045829; PubMed=8890751;  
RA Wang C., Entwistle J., Hou G., Li Q., Turley E.A.;  
RT "The characterization of a human RHAMM cDNA: conservation of the  
hyaluronan-binding domains."  
RL Gene 174:299-306(1996).  
CC -!- FUNCTION: INVOLVED IN CELL MOTILITY. WHEN HYALURONAN BINDS TO  
HMNR, THE PHOSPHORYLATION OF A NUMBER OF PROTEINS, INCLUDING THE  
FOCAL ADHESION KINASE OCCURS. MAY ALSO BE INVOLVED IN CELLULAR  
TRANSFORMATION AND METASTASIS FORMATION, AND IN REGULATING  
EXTRACELLULAR-REGULATED KINASE (ERK) ACTIVITY.  
CC -!- SUBUNIT: SUBUNIT OF THE HARC COMPLEX.  
CC -!- SUBCELLULAR LOCATION: CELL SURFACE AND CYTOPLASMIC (BY  
SIMILARITY).  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=A;  
CC IsoId=O75330-1; Sequence=Displayed;  
CC Name=B;  
CC IsoId=O75330-2; Sequence=VSP\_004286;  
CC -!- TISSUE SPECIFICITY: EXPRESSED IN BREAST CANCER CELL LINES AND IN  
NORMAL BREAST TISSUE.  
CC -!- DATABASE: NAME=PROW; NOTE=PROW 2.76-84(2001);  
CC WWW="http://www.ncbi.nlm.nih.gov/prow/guide/802868666.g.htm".  
-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AF032862; AAC32548.1; --  
EMBL; U29343; AAC52049.1; --  
Genew; HGNC:5012; HMWR.  
MIM; 60936;  
GO; GO:0007048; P:oncogenesis; TAS.  
KW Hyaluronic acid; Alternative splicing; Repeat; Glycoprotein; Antigen.  
FT DOMAIN 635 645  
FT HYALURONIC ACID-BINDING (POTENTIAL).  
FT DOVAIN 657 666  
FT N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 133 133  
FT N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 477 477  
FT N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 567 567  
FT N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 588 588  
FT Missing (in isoform B).  
FT VARSPLIC 75 90  
FT F/Tid-VSP 004286.  
FT CONFLICT 75 75  
FT K -> KK (IN REF. 2).  
FT CONFLICT 103 103  
FT S -> R (IN REF. 2).  
FT CONFLICT 277 277  
FT E -> D (IN REF. 2).  
FT CONFLICT 298 298  
FT K -> T (IN REF. 2).  
FT CONFLICT 322 322  
FT K -> E (IN REF. 2).  
FT CONFLICT 330 332  
FT QSR -> REH (IN REF. 2).  
SQ SEQUENCE 724 AA; 84031 MW; EA68AD6D2A626926 CRC64;  
Query Match 4.6%; Score 144; DB 1; Length 724;  
Best Local Similarity 20.0%; Pred. No. 0.18;  
Matches 112; Conservative 91; Mismatches 220; Indels 136; Gaps 21;  
34 KIVETSTNDITIGSQFQFKSIDELAARSAVETFKQFDLQKHIESIENADS- 92  
215 KLVSIKEKID---EKSETEKLEVEIEISCASDOVEKYKIDIALEENLKEKQDEILSL 271  
93 --KLESGNVLA-----ARNNHQPMSPRNNSVETTVTSQ 130  
272 KQLENIIVLSKQVEDLVNKKLEKEKEDHVRNREHNVLEMQNL----- 321  
131 PSQIEVPETSNPEGGMCMSCGLRKYIYANISDQAKL---MEEIPSAKLAKPAK 187  
322 -KQKFILEQOER-EKIQKQELQIDSLQKELSSHLQKLCFOEMVKEKNLFEELK 379  
188 FVLDCIGKPYL---QGRRAFTKSPSSARQVSLILLESFLMPDRGKVKLISWIKOE 244  
380 QTLDLCKLOQKEEQERLVKOLEEAKSRABELKLLBEKL-----KQK-----EAE 426  
245 AETAAVAKRRLMTEGLAAAEKMDARGLLLVACFGVPSNFRS----- 288  
427 LEKSSAA-----HTQATLLQEKYDS---MVQSLDVTAFESYKALTASEIEDLKLEN 477  
289 TDLLDIRMGSN-----EIALGKRSQFLVPMVSGIVESSIKGMHIALEWVYTFGM 342  
478 SSLQKAAKAGKNAEDVQHQILATESNSQYRYMLDLQTKSALKEIKETISVFLQKI 537  
343 EDKFSALVLTSLFKSKESPERAKRKAQSPAPFAEATKQLAVLSSVMQCWETHKLDPA 402  
538 TD-----LQNLKQOEEDFRKQLEDESGRAEKENTIAEL-----T 573  
403 KELPGWQ-IKEQIVSLKED-TLQLDK-EMEEKARSLSLMEEAALAKRYN----- 449  
574 EEINKWRLLYEELYNKTKPQLQDAFEVQKA---LILNEHGAQEQQLNKIRDSYAKLL 629  
450 --CQIKPEPLSPMEPPVTVSSSYSPYRDRSPFSQDDDDCDEISALVSSYLQ-----PST 502  
630 GHNLKQKIKHVKLKDENSQKSEVSKURCQLAKKQSQETKLQELNKLVLGIKHFDESK 689  
503 SFFHRSRSPYVPLPHG 521  
690 AFHHESENKALTKPLKEG 708

## RESULT 3

RB22 HUMAN STANDARD; PRT; 1722 AA.  
ID RB22 HUMAN STANDARD; PRT; 1722 AA.  
AC P29375;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Retinoblastoma-binding protein 2 (RBBP-2).  
GN RBBP2 OR RBP2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID:9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94020841; PubMed=8414517;  
RX Fattaey A.R., Helin K., Dembski M.S., Dyson N., Harlow E.,  
Vuocolo G.A., Hanobik M.G., Haskell K.M., Oliff A., Defeo-Jones D.,  
Jones R.E.;  
RA "Characterization of the retinoblastoma binding proteins RBP1 and  
RT RBP2.";  
RL Oncogene 8:3149-3156(1993).  
RN [2]  
RP SEQUENCE OF 1102-1569 FROM N.A.  
RX MEDLINE=91312450; PubMed=1857421;  
RX Defeo-Jones D., Huang P.S., Jones R.E., Haskell K.M., Vuocolo G.A.,  
Hanobik M.G., Huber H.E., Oliff A.;  
RA "Cloning of cDNAs for cellular proteins that bind to the  
RT retinoblastoma gene product.";  
RL Nature 352:251-254(1991).  
CC -!- FUNCTION: INTERACTS WITH THE VIRAL PROTEIN-BINDING DOMAIN OF THE  
CC RETINOBLASTOMA PROTEIN.  
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
CC -!- SIMILARITY: TO THE C-TERMINAL DOMAIN OF THE XE169 PROTEIN AND  
CC THE JUMONJI PROTEIN.  
CC -!- SIMILARITY: Contains 3 PHD-type zinc fingers.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; S66431; AAB28544.1; --  
PIR; I78879; I78879.  
Genew; HGNC:9886; RBBP2.  
MIM; 180202; --  
GO; GO:0005634; C:nucleus; TAS.  
GO; GO:0005515; F:protein binding activity; TAS.  
GO; GO:0003700; F:transcription factor activity; TAS.  
GO; GO:0003666; F:transcription from Pol II promoter; TAS.  
InterPro; IPR001606; ARID.  
InterPro; IPR003347; TF\_JmJc.  
InterPro; IPR003349; TF\_JmJN.  
InterPro; IPR004198; ZnF\_CSHC2.  
InterPro; IPR001965; ZnF\_PHD.  
Pfam; PF01388; ARID; 1.  
Pfam; PF02373; JmJc; 1.  
Pfam; PF02375; JmJN; 1.  
Pfam; PF00628; PHD; 3.  
Pfam; PF02928; ZF-CSHC2; 1.  
SMART; SM00501; BRIGHT; 1.  
SMART; SM00558; JmJc; 1.  
SMART; SM00545; JmJN; 1.  
SMART; SM00249; PHD; 3.  
PROSITE; PS01359; ZF\_PHD\_1; 3.  
PROSITE; PS00166; ZF\_PHD\_2; 3.  
Trans-acting factor; Nuclear  
ZN\_FING 293 343 PHD-TYPE 1.

NCBI\_TaxID=44689;

[1] SEQUENCE FROM N.A.

RR MEDLINE=87092266; PubMed=3540939;

RX Warrick H.M., de Lozanne A., Reinward L.A., Spudich J.A.;

RA "Conserved protein domains in a myosin heavy chain gene from

RT Dictyostelium discoideum.";

RT Proc. Natl. Acad. Sci. U.S.A. 83:9433-9437 (1986).

RL [2]

RR PHOSPHORYLATION SITES, AND MUTAGENESIS.

RC STRAIN=AX2;

RX MEDLINE=90353583; PubMed=2387408;

RA Lucke-Vielmeier D., Schleicher M., Grabatin B., Wippler J.,

RA Gerisch G.;

RA "Replacement of threonine residues by serine and alanine in a

RT phosphorylatable heavy chain fragment of Dictyostelium myosin II.";

RL FEBS Lett. 269:239-243 (1990).

RL [3]

RR PHOSPHORYLATION SITES.

RC MEDLINE=8811225; PubMed=2828113;

RX Wagie G., Noegel A., Scheel J., Gerisch G.;

RA "Phosphorylation of threonine residues on cloned fragments of the

RT Dictyostelium myosin heavy chain.";

RL FEBS Lett. 227:71-75 (1988).

RL [4]

RR X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-762.

RC MEDLINE=99345066; PubMed=7619795;

RX Fisher A.J., Smith C.A., Thoden J.B., Smith R., Sutoh K., Holden H.M.

RA Rayment I.;

RA "X-ray structures of the myosin motor domain of Dictyostelium

RT discoidium complexed with MgADP.Berx and MgADP.ALf4-.";

RL Biochemistry 34:8960-8972 (1995).

RL [5]

RR X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 1-762.

RC MEDLINE=95345067; PubMed=7619796;

RX Smith C.A., Rayment I.;

RA "X-ray structure of the magnesium (II)-pyrophosphate complex of the

RT truncated head of Dictyostelium discoidium myosin to 2.7-A

RT resolution.";

RL Biochemistry 34:8973-8981 (1995).

RL [6]

RR X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.

RC MEDLINE=96206189; PubMed=8611530;

RX Smith C.A., Rayment I.;

RA "X-ray structure of the magnesium (II).ADP.vanadate complex of the

RT Dictyostelium discoidium myosin motor domain to 1.9-A resolution.";

RL Biochemistry 35:5404-5417 (1996).

RL [7]

RR X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-762.

RC MEDLINE=97452580; PubMed=9305951;

RX Gulick A.M., Bauer C.B., Thoden J.B., Rayment I.;

RA "X-ray structures of the MgADP, MgATPgammA, and MgAMPNP complexes

RT of Dictyostelium discoidium myosin motor domain.";

RL Biochemistry 36:11619-11628 (1997).

RL [8]

RR X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.

RC MEDLINE=98070605; PubMed=9405148;

RX Bauer C.B., Kuhlman P.A., Bagshaw C.R., Rayment I.;

RA "X-ray crystal structure and solution fluorescence characterization

RT of Mg.2 (3')-O-(N-methylantraniloyl) nucleotides bound to the

RT Dictyostelium discoidium myosin motor domain.";

RL J. Mol. Biol. 274:394-407 (1997).

RL [9]

CC -1- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE

CC ACTIVITY THAT IS ACTIVATED BY ACTIN.

CC -1- SUBUNIT: MYOSIN II HEAVY CHAIN IS TWO-HEADED. IT SELF-ASSEMBLES

CC INTO FILAMENTS. HEXAMER OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI

CC LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS

CC (MLC-2).

CC -1- SUBCELLULAR LOCATION: HIGHEST CONCENTRATION IN THE POSTERIOR CELL

CC CORTEX.

CC -1- DOMAIN: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT

CC MYOSIN (1MM) AND 1 HEAVY MYOSIN (HMM). IT CAN BE FURTHER

CC SPLIT INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED

RESULT 4	MYS2_DICDI	STANDARD;	PRT; 2116 AA.
ID	MYS2_DICDI		
AC	P08799;		
DT	01-NOV-1988 (Rel. 09, Created)		
DT	01-OCT-1989 (Rel. 12, Last sequence update)		
DT	15-SEP-2003 (Rel. 42, Last annotation update)		
DE	Myosin II heavy chain, non muscle.		
GN	MYHCA.		
GN	MYHCA.		
OC	Dictyostelium discoideum (Slime mold).		
OC	Eukaryota: Mycetozoa: Dictyostellida: Dictyostelium.		

SUBFRAGMENT (82).

-1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES, CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.

-1- PTM: PHOSPHORYLATION INHIBITS THICK FILAMENT FORMATION AND REDUCES THE ACTIN-ACTIVATED ATPASE ACTIVITY.

-1- MISCELLANEOUS: DICTYOSTELIUM MYOSIN II HAS NO K(2)EDTA ATPASE ACTIVITY, PERHAPS CORRELATED WITH THE ABSENCE OF A CYS AT THE SH-1 POSITION (688).

-1- SIMILARITY: Contains 1 myosin-like globular head domain.

-1- SIMILARITY: Contains 1 IQ domain.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; M14628; AAA33227.1; --

FIR; A26655; A26655.

PDB; 1MMA; 03-DEC-97.

PDB; 1MMD; 17-AUG-96.

PDB; 1MMG; 03-DEC-97.

PDB; 1MMN; 03-DEC-97.

PDB; 1MND; 17-AUG-96.

PDB; 1MNE; 17-AUG-96.

PDB; 1VOM; 23-DEC-96.

PDB; 1LVK; 28-JAN-98.

PDB; 1DOX; 20-DEC-00.

PDB; 1DOY; 20-DEC-00.

PDB; 1DOZ; 20-DEC-00.

PDB; 1D1A; 20-DEC-00.

PDB; 1D1B; 20-DEC-00.

PDB; 1D1C; 20-DEC-00.

PDB; 1FMV; 20-DEC-00.

PDB; 1FMW; 20-DEC-00.

PDB; 1GBX; 17-JAN-01.

PDB; 1JMY; 07-NOV-01.

PDB; 1JX2; 07-NOV-01.

DictyDb; DD01008; mhca.

InterPro; IPR000048; IQ\_region.

InterPro; IPR001609; myosin\_head.

InterPro; IPR004009; Myosin\_N.

Pfam; PF00612; IQ; 2.

Pfam; PF00663; myosin\_head; 1.

Pfam; PF02736; Myosin\_N; 1.

PRINTS; PR00193; MYOSINHEAVY.

ProDom; PD000355; myosin\_head; 1.

SMART; SM00015; IQ; 1.

SMART; SM00242; MYSC; 1.

PROSITE; PS50096; IQ; 1.

Myosin; Coiled coil; Actin-binding; ATP-binding; 3D-structure; Calmodulin-binding; Methylation; Alkylation; Phosphorylation.

DOMAIN 1 761

DOMAIN 762 791

DOMAIN 817 2116

NP\_BIND 179 186

DOMAIN 638 660

DOMAIN 738 752

MOD\_RES 130 130

MOD\_RES 678 678

MOD\_RES 1823 1823

MOD\_RES 1833 1833

MOD\_RES 2029 2029

TURN 3 5

TURN 7 8

TURN 10 15

HELIX 25 28

HELIX 34 37

STRAND 41 42

TURN 48 55

FT STRAND 59 63

FT STRAND 69 73

FT HELIX 74 76

FT STRAND 78 79

FT HELIX 83 85

FT TURN 86 87

FT STRAND 90 90

FT STRAND 91 93

FT HELIX 99 110

FT TURN 111 113

FT STRAND 116 119

FT TURN 120 121

FT STRAND 122 126

FT STRAND 137 142

FT TURN 143 145

FT HELIX 148 150

FT HELIX 155 159

FT TURN 170 179

FT STRAND 173 179

FT TURN 181 182

FT HELIX 185 200

FT HELIX 210 226

FT STRAND 227 228

FT TURN 234 235

FT STRAND 240 247

FT TURN 249 250

FT STRAND 253 261

FT HELIX 265 268

FT TURN 269 289

FT TURN 273 274

FT STRAND 278 278

FT HELIX 279 287

FT HELIX 290 296

FT TURN 297 297

FT HELIX 301 303

FT TURN 305 307

FT TURN 316 317

FT HELIX 320 334

FT TURN 335 335

FT HELIX 338 355

FT TURN 356 356

FT STRAND 360 360

FT STRAND 368 368

FT HELIX 373 382

Query Match 4.3%; Score 134; DB 1; Length 2116;

Best Local Similarity 20.6%; Pred. No. 3.1;

Matches 123; Conservative 82; Mismatches 210; Indels 182; Gaps 27;

Qy 16 ANPLQHQHQEQRRRLPK-IVETESTMDITIGSKQPOFLKS-----IDELAPSVAV 69

Db 817 ARPLKRRNFKEIKEREILELKSNTLDTSTQDKLEKSKDKTESNVLQOLKAEK 876

Qy 70 EYFKRQFD-----DLQGHIESIENAIKSLGNSGVVLAARN----- 105

Db 877 ETLKMYDSKDALEAKRELEIRVEDMESELDEK-----KALENLQNKQKRSVEEKVRD 930

Qy 106 -----NNFHQPMLSFPRNNVSVETVT-VSQPQOEIVPETSNTPEG 145

Db 931 LEELEQEQKRLNTLEKLKKYEELEEEMKRVNDQSDTISRLEKIKDELQKVEELETES 990

Qy 146 GRMCELMCSKGL---RKVIYANISD-----OAKLMEEI----- 175

Db 991 --FSESKDKGVLEKTRVRLQSELDDLTVRLDSETKOKSELLROKKLEBELKQVQAL 1048

Qy 176 -PSALKLAKPAKFLVDCIGKPYLQGRRAFTKESPMSSARQVSLILLESFLL-----MPD 229

Db 1049 AETAAKLAQEAANKKLQ--GE-YTELNEKFNSEVTARSNVEKSKTLESQLVAVNNELDE 1105

Qy 230 RGKGVKIESWIKDEAETAATAVAWARK-RLWTEGG-----LAAAEKMDARGLLLVACFG 281

Db 1106 EKNRDALK--KKKALDAMLEEMKDKQLESTGGEEKKSLYDLKVKQESDMELRNQI----- 1159

```
QY 282 VPSNFRST-DLLDLIRMSGNEIAGALKRSQFLVPMVSGIVSESSIKRMHIEALEMVTFF 340
DB 1160 --SELQSTTIALEKIKSTLEGEVA-----RLQGELEAEQLAKSNVYKQKKVEL 1206
QY 341 GHEDFSAALVLTSLFKMSKESFERAKAKAOSPLA-----FKEAATK----- 382
DB 1207 DLEDK-SQALET--AAQALDKLKKLQELSEVOTQSEANNKVNDSSTNKHLET 1262
QY 383 -----QLAVLSVWQCMETHKLDPAKELPGWQIKQI-----VSLKOT 421
DB 1263 SFNNLKLAEARQAKALEKKRLGLESELK--HYNEQLSEEEKQKESNEKKRVDLEKEV 1320
QY 422 LOLDKMEBEKARSLISMEBAALAKRMVQIKRPLSPMEPPVTSSSYSPIDRDS 478
DB 1321 SELKQIEBEVASKAVTEAKNKESELDLKR-----QYADVSS-----RDXS 1365

RESULT 5
ID SCPI MOUSE STANDARD; PRT; 993 AA.
AC Q62259; O09205; P70192; Q62329;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Synaptonemal complex protein 1 (SCP-1 protein).
GN SYP1 OR SCPI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=CBA; TISSUE=Testis;
RX MEDLINE=96004899; PubMed=7548215;
RA Sage J., Martin L., Cuzin F., Rassoulzadegan M.;
RT "cDNA sequence of the murine synaptonemal complex protein 1 (SCP1).";
RL Biochim. Biophys. Acta 1263:258-260(1995).
RN [2]
SEQUENCE FROM N.A.
RP STRAIN=Swiss; TISSUE=Testis;
RA Kerr S.M., Taggart M.H., Lee M., Cooke H.J.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
SEQUENCE OF 1-149 FROM N.A.
RP STRAIN=C57BL/6;
RA Sage J., Li Y., Martin L., Mattei M.-G., Guenet J.-L., Liu J.G.,
RA Hoog C., Cuzin F., Rassoulzadegan M.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
SEQUENCE OF 95-787 FROM N.A.
RP STRAIN=ICR; TISSUE=Testis;
RA Teuchida J., Nishina Y., Nozaki K., Nishimune Y.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAJOR COMPONENT OF THE TRANSVERSE FILAMENTS OF
CC SYNAPTONEMAL COMPLEXES (SCS), FORMED BETWEEN HOMOLOGOUS
CC CHROMOSOMES DURING MEIOTIC PROPHASE.
CC -!- SUBCELLULAR LOCATION: NUCLEAR. IN TRIPARTITE SEGMENTS OF
CC SYNAPTONEMAL COMPLEXES, BETWEEN LATERAL ELEMENTS IN THE NUCLEUS.
CC FOUND ONLY WHERE THE CHROMOSOME CORES ARE SYNAPSED. ITS N-TERMINUS
CC IS FOUND TOWARDS THE CENTRE OF THE SYNAPTONEMAL COMPLEX WHILE THE
CC C-TERMINUS EXTENDS WELL INTO THE LATERAL DOMAIN OF THE
CC SYNAPTONEMAL COMPLEX (BY SIMILARITY).
CC -!- DOMAIN: CONSISTS OF AN ALPHA-HELICAL STRETCH OF 700 AA RESIDUES,
CC FLANKED BY N- AND C-TERMINAL GLOBULAR DOMAINS. THE C-TERMINAL
CC DOMAIN HAS DNA-BINDING CAPACITY (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/)
```

```
CC or send an email to license@isb-sib.ch).
CC -----
DR Z38118; CRA86262.1; -
DR EMBL; L41069; AAC64514.1; ALT INIT.
DR EMBL; U62864; AAC53335.1; -
DR EMBL; U62860; AAC53335.1; JOINED.
DR EMBL; U62861; AAC53335.1; JOINED.
DR EMBL; U62862; AAC53335.1; JOINED.
DR EMBL; U62863; AAC53335.1; JOINED.
DR EMBL; D88539; BAAI3639.1; -
DR PIR; S49461; S49461.
DR MGD; MGI:105931; Sycp1.
KW Nuclear protein; Meiosis; Cell division; Phosphorylation;
KW DNA-binding; Coiled coil.
FT DOMAIN 12 97
FT DOMAIN 104 815
FT DOMAIN 114 117
FT DOMAIN 597 700
FT DOMAIN 898 901
FT DOMAIN 978 986
FT CONFLICT 527 527
FT SEQUENCE 993 AA; 115962 MW; 1A4FA790D64FAFE6 CRC64;
Query Match 4.2%; Score 131; DB 1; Length 993;
Best Local Similarity 19.8%; Pred. No. 1.7;
Matches 122; Conservative 97; Mismatches 222; Indels 176; Gaps 29;
QY 13 TTTANPLLRHQSQRRELPLKIVETESTMDITIGOSKOPQLKSIDELAAF----SVA 68
DB 381 TCTLEELL---RTEQORLE-----KNEDQLKLITVELQKS---NELEWTKFNKKEVE 429
QY 69 VETPK-----QFDDLQKHIESIENAIIDSKLSNGVVLAAARNNHQPMSPRNVSV 122
DB 430 LEELKNILAEQDKLLDEKKQVLAELQEQELFTLLETRKEVH-----DLQE 480
QY 123 ETTVT-----VSQSQEIVPETSNNKPEGRMCMLCCKGLAKYIYANISDOAKLM 172
DB 481 QVTVTISQHYLKQVEEMKTELEKLNLTETASCDML-----LLENKFEV 528
QY 173 EETPSALKLAKEPAKFLVDCIGFYLOGRRAFYKESPMSSARQVSLILLESFLMLPDRGK 232
DB 529 QEASDMALELKHQEDINC-----KKQERLLKQIENL----- 562
QY 233 GKVKTIESWIKDEAETAAVANRKLMTGG-----LAAAEKMDARGL-----LL 275
DB 563 --EEKEMHRLDELSV-----RKEFIQGGDEVCKLDKSEE--NARSICEVLEKKEKMKIL 615
QY 276 LVACFGVPSNF--RSTDLDL-----IRMSGNEIAGALKRSQFLVPMVSGIVSESSIKR 327
DB 616 ESKCNLKKQVENKSKNIEELHQENKTLKKKSAEIK-QLNAYEIKVSKLELESTKQR 674
QY 328 GNHIEALEMVTFFGMEK-----FSAALVLTSLFKMSKESFERAKRAQSPFLAF 376
DB 675 ---FEEMTNNYQKBIENKKISEGKLGEVEKAKATDEAVKLOKEIDLRCQHKIAEWAL 731
QY 377 KEAAATKQ-----LAVLSVMQCMETHKLDPAKELPGWQIKQIEIVSLEKDTLQD 425
DB 732 MEKHKHQYDKIVERDSSELGLYKNRECEQSSAKIALETLS--NIRNELVSLK--QLE 786
QY 426 KEMBEKARSLIMEBAALAKRMVQIKRPLSPMEPPVTSSSYPIYDRSFPQR--- 483
DB 787 IEKEEKEKLMKAKENTAILK---DKDKTKIQLSLESPEATSKWFD---SKTTPSCNIS 839
QY 484 -----DDQDETALVSSYLGPTSFPHRRRSPEYVPLPHGGLGRSVYAEH 532
DB 840 RLSSMSGSKSKDNRDLRASAKIL-PTTV-----TKEYTVKTP---TKSKIYQRE- 887
QY 533 LAPNSYSPGHRLHRQ 549
DB 888 ---NKYIPTGGSNKKRK 901
RESULT 6
```

SCPI1 RAT  
ID -SCPI1 RAT STANDARD; PRT; 997 AA.  
AC 003410;  
DT 01-JUN-1994 (Rel. 22, Created)  
DT 15-JUN-1998 (Rel. 36, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE Synaptonemal complex protein 1 (SCP-1 protein).  
GN SYCP1 OR SCPI1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RX MEDLINE=93099894; PubMed=1464329;  
RA Meuwissen R.L.J., Offenbergh H.H., Dietrich A.J., Riesenwijk A.,  
RA Iersel M., Heyting C.,  
RT "A coiled-coil related protein specific for synapsed regions of  
RT meiotic prophase chromosomes.";  
RL MBO J. 11:5091-5100(1992).  
CC -1- FUNCTION: MAJOR COMPONENT OF THE TRANSVERSE FILAMENTS OF  
CC SYNAPTONEMAL COMPLEXES (SCS), FORMED BETWEEN HOMOLOGOUS  
CC CHROMOSOMES DURING MEIOTIC PROPHASE.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR. IN TRIPARTITE SEGMENTS OF  
CC SYNAPTONEMAL COMPLEXES, BETWEEN LATERAL ELEMENTS IN THE NUCLEUS.  
CC FOUND ONLY WHERE THE CHROMOSOME CORES ARE SYNAPSED. ITS N-TERMINUS  
CC IS FOUND TOWARDS THE CENTRE OF THE SYNAPTONEMAL COMPLEX WHILE THE  
CC C-TERMINUS EXTENDS WELL INTO THE LATERAL DOMAIN OF THE  
CC SYNAPTONEMAL COMPLEX.  
CC -1- TISSUE SPECIFICITY: Testis.  
CC -1- DEVELOPMENTAL STAGE: EXPRESSED EXCLUSIVELY IN MEIOTIC PROPHASE  
CC CELLS.  
CC -1- DOMAIN: CONSISTS OF AN ALPHA-HELICAL STRETCH OF 700 AA RESIDUES,  
CC FLANKED BY N- AND C-TERMINAL GLOBULAR DOMAINS. THE C-TERMINAL  
CC DOMAIN HAS DNA-BINDING CAPACITY.  
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT WAS  
CC CORRECTED IN POSITION 6 TO MAXIMIZE THE SIMILARITY WITH THE  
CC OTHER SPECIES SYCP1 SEQUENCES.  
CC  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL; X67805; CA48006.1; ALT FRAME.  
CC Nuclear protein; Meiosis; Cell division; Phosphorylation;  
CC DNA-binding; Coiled coil.  
CC FT DOMAIN 108 819 COILED COIL (POTENTIAL).  
CC FT DOMAIN 118 121 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
CC FT DOMAIN 701 704 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
CC FT DOMAIN 902 905 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
CC FT DOMAIN 982 990 ARG/LYS-RICH (BASIC).  
CC SQ SEQUENCE 997 AA; 116511 MW; 229D59823FD684BE CRC64;  
Query Match 4.2%; Score 129; DB 1; Length 997;  
Best Local Similarity 19.1%; Pred. No. 2.3; Indels 146; Gaps 25;  
Matches 115; Conservative 106; Mismatches 235;  
13 TTANPLLOHQSEQRRLPKIVETESTSDITIGQSKQPOFLKSIDELAAAF----SVA 68  
385 TCTLEELL---RTEQQRLE-----NNEQDLKUITMELQKS---SELEMTKFNKKEVE 433  
69 VETFK-----RQFDLQKHISIEAIDSKLESNGVLAARNNNFHQPMLSPPRNVSU 122  
434 LEEKLTILADQKLLKQVKEKLAELQKQKQELTFLLOTREKIH-----DL 482  
123 ETTVTVSQSQE----IVPETSNNKPGGGM---CELMCSKGLKTYIYANISDQAKIMEIP 176  
483 EVQVTVTKTSEEHYKQVEMKTELEKXKNIELTANSDML-----LLENKKLVQEQAS 536

177 SALKLAKPAKFVLDICIGFYIQQRAFTKESPMSSARQVSLILLESFLLMPDRGKGVK 236  
537 DMVLEKKGQEDIIINCKQBERMLAQIETLEEKWNLDELESVRKEFIQOGDEVKCKL- 595  
237 IESWIKDEAETAAVAWRKRLMTEGGLAAAKODARGULLLVACFGVPSNF--RSTDLIDL 294  
596 -----DKSEENARSTIEVL-----KKEQMKILENCKNNLKQIENKKNIEEL 640  
295 -----IRVSGSN-----BIAGALKESQFLVPMVSGIVSESSIKRGVHI 331  
641 HOENKALKKSSAENKQNAVEIKVKNKLELSTKQKFEEMI-----NNTQKEIEI 692  
332 EALEMVTYTFMEDKFSALVLTSLFMGSKESFERAKRAQSPFLAFKEAATKQ----- 383  
693 KKISEBKLLGEVEK--AKATVDEAVKLQKEIDLRCOHKIAEWALMEKHQYDKIVEER 750  
384 ---LAVLSSVMQCMETHKLDPAKELPGHQIKEQIVLSLEKOTLDKEMEEKARLSLME 440  
751 DSELGLYKNREQEQSSAKVALETELS--NIRNELVSLKK--OLEVEKEBEKJLMQEEN 805  
441 AALAKRMYNQOIKRPLSPMEMPPVTSSSYSPYVDRSRFPSPQR-----DDQ 487  
806 TAI---LTDKDKKIQASLLESPEATSNKFD-----SKTTPSQNISRLLSSMDSGKSDNR 858  
488 DEISALVSSYLGPSTSPFHRGRRSPYVPLPHGGLGRSVYAYEHLAPNSYSPGHGRHLH 547  
859 DSLRASAKSILSTTV-----TKEYTVKTP---TKKSIYQRE---NKYLP7GGSNKK 903  
548 RQ 549  
904 RK 905  
RESULT 7  
KI67 HUMAN  
ID KI67 HUMAN STANDARD; PRT; 3256 AA.  
AC P46013;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Antigen KI-67.  
GN MKI67.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94043435; PubMed=8227122;  
RA Schluter C., Duchrow M., Wohlenberg C., Becker M.H.G., Key G.,  
RA Flad H.-D., Gardes J.;  
RT "The cell proliferation-associated antigen of antibody Ki-67: a very  
RT large, ubiquitous nuclear protein with numerous repeated elements,  
RT representing a new kind of cell cycle-maintaining proteins.";  
RL J. Cell Biol. 123:513-522(1993).  
RN [2]  
RP SEQUENCE OF 1-31 FROM N.A.  
RA Gerdes J.;  
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: THOUGHT TO BE REQUIRED FOR MAINTAINING CELL  
CC PROLIFERATION.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR. PREDOMINANTLY LOCALIZED IN THE G1  
CC PHASE IN THE PERINUCLEAR REGION, IN THE LATER PHASES IT IS ALSO  
CC DETECTED THROUGHOUT THE NUCLEAR INTERIOR, BEING PREDOMINANTLY  
CC LOCALIZED IN THE NUCLEAR MATRIX. IN MITOSIS, IT IS PRESENT ON ALL  
CC CHROMOSOMES.  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=Long;  
CC IsoID=P46013-1; Sequence=Displayed;  
CC Name=Short;  
CC IsoID=P46013-2; Sequence=VSP\_004298;

-!- DEVELOPMENTAL STAGE: EXPRESSION OF THIS ANTIGEN OCCURS  
PREFERENTIALLY DURING LATE G1, S, G2, AND M PHASES OF THE CELL  
CYCLE. WHILE IN CELLS IN G0 PHASE THE ANTIGEN CANNOT BE DETECTED.  
-!- SIMILARITY: Contains 1 FHA domain.  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
EMBL; X65550; CAA46519.1; -  
EMBL; X65551; CAA46520.1; -  
EMBL; X94762; CAA64388.1; -  
PIR; A48666; A48666.  
Genew; HGNC:7107; MKI67.  
GK; P46013; -  
MIM; 176741; -  
GO; GO:000074; P:regulation of cell cycle; TAS.  
InterPro; IPR000253; FHA.  
Pfam; PF00498; FHA; 1.  
SMART; SM00240; FHA; 1.  
PROSITE; PS0006; FHA DOMAIN; 1.  
Cell cycle; Antigen; Nuclear protein; ATP-binding; Repeat;  
Alternative splicing; Polymorphism.  
DOMAIN 27 76  
FT DOMAIN 1000 2928 16 X 122 AA APPROXIMATE REPEATS.  
FT REPEAT 1000 1112 1.  
FT REPEAT 1122 1234 2.  
FT REPEAT 1244 1356 3.  
FT REPEAT 1366 1477 4.  
FT REPEAT 1487 1598 5.  
FT REPEAT 1608 1720 6.  
FT REPEAT 1730 1842 7.  
FT REPEAT 1851 1964 8.  
FT REPEAT 1974 2086 9.  
FT REPEAT 2096 2204 10.  
FT REPEAT 2214 2326 11.  
FT REPEAT 2335 2447 12.  
FT REPEAT 2457 2569 13.  
FT REPEAT 2579 2689 14.  
FT REPEAT 2699 2808 15.  
FT REPEAT 2818 2928 16.  
FT NP BIND 3034 3041 ATP (POTENTIAL).  
FT VARSPLIC 136 495 Missing (in isoform Short).  
FT VARIANT 3150 3150 /FTid=VSP\_004298.  
FT VARIANT 3217 3217 T->S (IN dSNP:11106).  
FT VARIANT 3217 3217 K->E (IN dSNP:8473).  
FT SEQUENCE 3256 AA; 358741 MW; 578P8C51BBD42517 CRC64;  
Query Match 4.1%; Score 127.5; DB 1; Length 3256;  
Best Local Similarity 20.0%; Pred. No. 14;  
Matches 137; Conservative 92; Mismatches 228; Indels 229; Gaps 32;  
QY 11 OPTTANPLQRHQSEQR--RELPKIVETESTMDI-----TIGQSKQPOFLK 57  
DB 864 EPSKTVSTVNRSGRSTPRNIQKLPVSKSEBTETVEICILKRGQKATLLQORREGENK 923  
QY 58 SDELAARSVAVETKRFQDLDLKHIEIENALDKLESNGVLAARNPHQ----PMLS 114  
DB 924 EIE-----RPFETYNKELKEN--DEKMK-----MKRSRTWGQKCAPM-- 961  
QY 115 PPRNNVSVETTVTSQSQEIVPETS-----KPEGGRMCLMCSKGLRKY 160  
DB 962 -----SDLTDLKSLPOTELMKDTARGQNLLQTDHAKAPKSEKGIKMPQCSLQPEP 1014  
QY 161 IVANISDQAKLMEETPSALKAKEPAKVLDCIGKF-VLQRRRAFTVESPMSARQSLL 219  
DB 1015 INTPHTKQQL-----KASLGKGVKGBELLAVGKFTRTSGTTHTRHPAGCKSI--- 1065

RESULT 8

PCT2 MOUSE  
ID PCT2 MOUSE STANDARD; PRT; 1920 AA.  
AC P48725;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Pericentrin 2.  
GN PCNT2 OR PCNT.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94170365; PubMed=8124707;  
RA Dosey S.J., Stein P., Evans L., Calarco P.D., Kirschner M.;  
RT "Pericentrin, a highly conserved centrosome protein involved in  
microtubule organization."  
RL Cell 76:639-650(1994).  
CC -!- FUNCTION: INTEGRAL COMPONENT OF THE FILAMENTOUS MATRIX OF THE  
CENTROSOME INVOLVED IN THE INITIAL ESTABLISHMENT OF ORGANIZED  
MICROTUBULE ARRAYS IN BOTH MITOSIS AND MEIOSIS.  
CC -!- TISSUE SPECIFICITY: HIGHEST LEVELS FOUND IN KIDNEY, THYMUS AND  
LIVER. LOW LEVELS IN BRAIN, MUSCLE, LUNG AND HEART.  
CC -!- DOMAIN: COMPOSED OF A COILED-COIL CENTRAL REGION FLANKED BY NON-  
HELICAL N- AND C-TERMINALS.  
CC -!- SIMILARITY: STRONG TO HUMAN KENDRIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
EMBL; U05823; AAA17886.1; -  
PIR; A53188; A53188.

QY 220 ILBSFLMPD-----RGKKYKIESWIKDEATAA-VAMKRLMTTEGGAAAEKMDAR 271  
DB 1066 --RTFKSPQILDPAARVTGMKKWPRTPKEAQSLDELGKFKLFTQPGPSESWTDEK 1123  
QY 272 GLILLVACFVP-----SNFRSTDL-----LDLRMSGNEIAGALKRSQF 312  
DB 1124 --TTKIACKPPSPESVDPTFTQWPKRSIRKADVEEFALRKUTPS--AG-----KAM 1174  
QY 313 LVPVSGIVESSIK--RGMHIEALEMVTYFGMEDFSAALVLTSLFKMSKESFEERAKKA 370  
DB 1175 LTPKAGGDEKDKAFMGFTPVQKLDLAGT-----LPSKRLQLOTPKEKA 1218  
QY 371 QS--PLA-FYE-----AATKO----- 383  
DB 1219 QALEDLAGFKELFQTPGHTTEELVAAGTKTKIPCDSPQSDPVPTFTSTKQPKRSIRKADV 1278  
QY 384 ----LA---VLSVMQCMET-----HKLDPAKELPGWQIKQIVSL 417  
DB 1279 EGGLEACENLMPAGRAMHTPKPSVGEKXDIIFVGTVPQKLDLTENLTGSKRRPTPK 1338  
QY 418 EKOTQLDKMEWEKARSLSLMEEAALKKMYNQOIKRPLSPMEVPPVTSYSSYIYRDR 477  
DB 1339 EAQALEDLTGKELFQTPGHTTEEAAG-----KTKMPCESSFPESADFTTTRQP 1391  
QY 478 SFPQRDDDDDETSAL--VSSYLGPTSTFPHRSRRSPPEYVPLPHGLGRSVVAYEHLAP 535  
DB 1392 KTFPEKRDVQKLSALKKLTQTSGET--HTDK-----VP---GGEDKSINAFREIAK 1439  
QY 536 NSYSPGHGHLHRQYSPSLVHGQRHP 561  
DB 1440 QKLDP-----AASVTGSKRHP 1455

DR MGD; MGI:102722; PcnL2.  
DR GO; GO:0005813; C:centrosome; IDA.  
KW Coiled coil; Microtubules.  
FT DOMAIN 110 1600  
FT DOMAIN 109 112  
FT DOMAIN 383 387  
SQ SEQUENCE 1920 AA; 218337 MW; CF1D0ADCE5B73309 CRC64;  
  
Query Match 4.1%; Score 126.5; DB 1; Length 1920;  
Best Local Similarity 19.8%; Pred. No. 7.8; Indels 141; Gaps 25;  
Matches 105; Conservative 99; Mismatches 186; Indels 141; Gaps 25;  
  
QY 21 QRHSQRRELFKIVETSTMDITIG---QSKQPFKSIDELAAPSVA-VETFKRQF 76  
DB 679 QLASSESRQAL--LETHVAEQVKNHAEISALEKRLHSLNLEDESCYADVQVIR--- 732  
QY 77 DLQKHIEIENAIKSLNSGVLAAANNPHQMLSPRRNV-----SVETVT 127  
DB 733 DEHQALELLRAELBEQLQ-----KKECHREMLTQELLENKRAHAEQLQVRSLR 784  
QY 128 VSQPSQEIIV-----PETSINKPEGGRMCE-----LMCSKGLR 158  
DB 785 MENSAGHIENGKPAADLOGAHQDQPPAMALHNEGHLLVEDGDVILRSVDAEGLLHQAGPQ 844  
QY 159 KYIYANISQAKMEI--PSALKIAKEPAKFVLCIGKFIQGR-----PAFTKESFMS 211  
DB 845 ELGDAHTVEMQKSAELAKPQELQASQDQVAVRD---KVFLNLELECECRADVEQLQQR 901  
QY 212 SARQ-----VSLILFELFLLMPGRGKVKIESWIKD-----EAETAAVARKRLM 257  
DB 902 RERENEGTTLICMLRADLELAQGEKA-----LDAALRRULLDFGDTLKAAYTLKSEIS 956  
QY 258 TEGGLA-----AAKMDARGLLLLLVACFG-----VPSNFRSTDLIDL 294  
DB 957 EGAGLLLDHEDADTSDAR---LAAAPLGDMMWSDGLLEIDRTLPGEAETSVCESISHV 1013  
QY 295 -----IRMSGNEIAGALKR-SQFLVPMVSGIVESSIKKGWHIEALEMVTYTFGMEDKPSA 348  
DB 1014 CESFFRPNETLDCQPIRRVYQSLSTAVEGLEHVALDSSKOLEARQLHRC-VEREF-- 1070  
QY 349 ALVLTSLFKMSKESFBRARKAQSPL-----AFKEAATKOLAYLSSVMQCMETHKLDP 401  
DB 1071 -----RHRNEBMAQAMQKQELLERLREBSAARDRLALEHTAKGLLEGFKVEKYVDL 1122  
QY 402 AKELPGWQIKQEVLSLEKTLQDKEMEEKARS--LSLMEEAALAKEMYNQO 451  
DB 1123 QEALGKKESEQOQILELEDLR--KOLEQAARELUTLKEEKSV---LWNQK 1168  
  
RESULT 9  
SR72\_CAEEL STANDARD; PRT; 694 AA.  
AC P91240;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Signal recognition particle 72 kDa protein homolog (SRP72).  
GN F08D12.1  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN SEQUENCE FROM N.A.  
RP STPAIN-Bristol N2;  
RC Le T.T., Waterston R.;  
RA Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
RL FUNCTION: SIGNAL-RECOGNITION-PARTICLE ASSEMBLY HAS A CRUCIAL ROLE  
CC IN TARGETING SECRETORY PROTEINS TO THE ROUGH ENDOPLASMIC  
CC RETICULUM MEMBRANE. SRP72 BINDS THE 7S RNA ONLY IN PRESENCE OF  
CC SRP68. THIS RIBONUCLEOPROTEIN COMPLEX MIGHT INTERACT DIRECTLY WITH  
CC THE DOCKING PROTEIN IN THE ER MEMBRANE AND POSSIBLY PARTICIPATE  
CC IN THE ELONGATION ARREST FUNCTION (BY SIMILARITY).

CC -!- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A 7S RNA MOLECULE  
CC OF 300 NUCLEOTIDES AND SIX PROTEIN SUBUNITS: SRP72, SRP68, SRP54,  
CC SRP19, SRP14 AND SRP9 (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: BELONGS TO THE SRP72 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U80840; AAB37925.1; -;  
CC F1R; T25685;  
CC Wormpep; F08D12.1; CE09234.  
CC InterPro; IPR001440; TPR.  
CC Pfam; PF00515; TPR; 3  
KW Signal recognition particle; Ribonucleoprotein.  
FT DOMAIN 684 689  
SQ SEQUENCE 694 AA; 78485 MW; BE12B69DC8B7C4FA CRC64;  
  
Query Match 4.1%; Score 126; DB 1; Length 694;  
Best Local Similarity 18.9%; Pred. No. 2.1;  
Matches 111; Conservative 92; Mismatches 199; Indels 186; Gaps 24;  
  
QY 35 IYETESTND-----ITIGOSQPFKSIDELAAPSVAVETFKRQFDDQLQKHISIEIEN 88  
DB 164 VAETEDSYSLYNACVIEAEKLPQALESEK-----ALKTKRKFEDREDEIEIE 217  
QY 89 AIDSKLESNGVGLAARNNN-----FHQMLSPRRNVSVETTVTSQP-----SQEIVPETS 140  
DB 218 ELDSTRVQKAYVQLRMQOKAEALAIYKVAQANHPDSVKATITNNIPASSDFALPESR 277  
QY 141 NKPEGG-RMCELMCSK-----GLRKYIYANISDQAKMEIIPSA 178  
DB 278 KRFAALQIFQSKCKSPDFWLKMLQSKSKTKFLQKXGNFTYFTSKNSQIYRFSQAIPPK 337  
QY 179 LKLAKEPAKFVLDICIGKFIQGRRAFTKESYWSARQVSLILLESFLLMPDRGKGVKIE 238  
DB 338 -----IPNFSFO-----IDTKLTRQRUTLMMNALVLL----- 367  
QY 239 SWIKDEAETAAVAMKRLMTGLEGLAAAEKMDARGLLLLLVACFG-----VPSNFRST 289  
DB 368 --LSNQREPC-----KRALEE-----LVAKFGSSKDVALLTEATLHFKMG 404  
QY 290 DLDLIRMSGNEIAGALKRSQFLVPMVSGIVESSIKKGWHIEALEMVTYTFGMEDKPSAA 349  
DB 405 DAEAAKVLGSDLEQSLARHLV-----LNAGRLPEAVGAIRDLFISGKLGAAS 453  
QY 350 LVLTSLF-----KMSKESPERAKRKAQSPLAFKE-----AATKOLAV 386  
DB 454 SLTSTLIAADSRDEAVKELVAASAKNQTPKALKSILEDLVEVEQOQGNETAATKHEK 513  
QY 387 L-----SSVMQCM---ETHKLDPAKELPGWQIKQEVLSLEKTLQDKEMEE----- 430  
DB 514 LVEKFPEDLQQLCRVLGVASKTDPKK-----ABSLSAKLPFETMEVDVNVNDELESDWI 567  
QY 431 -----KARSLMEBAALAKRMYNQOIKRPLSPMEMPPVTSSSYSP-----IY 474  
DB 568 LYGEKTRQKKEAKSPQTAETIAATRK-----LKATYKRRKIRLPKNYNSAVTDPERWLP 623  
QY 475 RDRS-FPSQRDQDDEI--SALVSSYLGPSTSFPHRSRRSPPEYMWPLP 519  
DB 624 QERSTYKRRKKNREIREIGRGTCGSSSANPNVEYVTASPNSPR---PLP 668  
  
RESULT 10  
GOG4\_HUMAN STANDARD; PRT; 2230 AA.  
ID GOG4\_HUMAN  
AC Q13439; Q13270; Q13654; Q14436;  
DT 16-OCT-2001 (Rel. 40, Created)

16-OCT-2001 (Rel. 40, Last sequence update)  
15-SEP-2003 (Rel. 42, Last annotation update)  
Golgi autotransfer, golgin subfamily A 4 (Trans-Golgi p230) (256 kDa golgin) (Golgin-245) (72.1 protein).  
GOLGA4.  
Homo sapiens (Human).  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NCBI\_TaxID=9606;  
[1]  
SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
MEDLINE=96215236; PubMed=8626529;  
Erich R., Gleeson P.A., Campbell P., Dietzsch E., Toh B.-H.;  
"Molecular characterization of trans-Golgi p230: a human peripheral membrane protein encoded by a gene on chromosome 6p12-22 contains extensive coiled-coil alpha-helical domains and a granin motif.";  
J. Biol. Chem. 271:8328-8337(1996).  
[2]  
SEQUENCE FROM N.A.  
Seelig H.P.;  
Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.  
[3]  
SEQUENCE OF 131-2230 FROM N.A.  
TISSUE=Placenta;  
MEDLINE=96125112; PubMed=8537393;  
Fritzler M.J., Lung C.-C., Hamel J.C., Griffith K.J., Chan E.K.L.;  
"Molecular characterization of golgin-245, a novel Golgi complex protein containing a granin signature.";  
J. Biol. Chem. 270:31262-31268(1995).  
[4]  
SEQUENCE OF 524-672 FROM N.A.  
TISSUE=Gastric fundus;  
Balague C.;  
Thesis (1994), Instituto municipal de investigacion medica, Spain.  
-!- FUNCTION: May play a role in vesicular transport from the trans-Golgi.  
-!- SUBCELLULAR LOCATION: ASSOCIATED WITH THE CYTOPLASMIC FACE OF THE GOLGI MEMBRANE.  
-!- ALTERNATIVE PRODUCTS:  
Event=Alternative splicing; Named isoforms=4;  
Comment=Additional isoforms seem to exist;  
Name=1;  
IsoId=Q13439-1; Sequence=Displayed;  
Name=2;  
IsoId=Q13439-2; Sequence=VSP\_004272, VSP\_004273;  
Name=3;  
IsoId=Q13439-3; Sequence=VSP\_004274;  
Name=4;  
IsoId=Q13439-4; Sequence=VSP\_004275;  
-!- DISEASE: ANTIGEN IN THE AUTOIMMUNE DISEASE SJOEGREN'S SYNDROME AND IN HEPATITIS B.  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
EMBL; U41740; AAC50434.1; -;  
EMBL; X82834; CAA58041.1; -;  
EMBL; U31906; AAC51791.1; -;  
EMBL; X76942; CAA54261.1; -;  
Genew; HGNC:4427; GOLGA4.  
MIM; 602509; -;  
GC; GO:0005802; C:Golgi trans-face; TAS.  
GC; GO:0016192; P:vesicle-mediated transport; TAS.  
InterPro; IPR000237; GRIP\_domain.  
Pfam; PF01465; GRIP; 1.  
Golgi stack; Antigen; Coiled coil; Alternative splicing.  
COILED COIL (POTENTIAL).  
DOMAIN 133 237  
DOMAIN 276 1011  
COILED COIL (POTENTIAL).

FT	DOMAIN	1033	1214	COILED COIL (POTENTIAL).
FT	DOMAIN	1259	2152	COILED COIL (POTENTIAL).
FT	VARSPPLIC	2154	2185	TPKGNLHIDVSLFGFTFEYLRKVLFEY --> HLTKV AICTRMSSLENPLNSICEKCLFSI (in isoform 2).
FT	VARSPPLIC	2103	2109	Missing (in isoform 3).
FT	VARSPPLIC	2222	2230	FTSPRSGLF --> SWIRSSS (in isoform 4).
FT	CONFLICT	188	188	FTID=VSP_004272.
FT	CONFLICT	220	220	FTID=VSP_004275.
FT	CONFLICT	276	276	R -> K (IN REF. 3).
FT	CONFLICT	584	584	Y -> H (IN REF. 3).
FT	CONFLICT	628	628	T -> A (IN REF. 3).
FT	CONFLICT	628	628	K -> E (IN REF. 3).
FT	CONFLICT	630	630	T -> A (IN REF. 3).
FT	CONFLICT	682	682	K -> E (IN REF. 3).
SQ	SEQUENCE	2230	AA; 261139	MW; 3BBY33DB1EA86134 CRC64; Query Match 4.1%; Score 126; DB 1; Length 2230; Best Local Similarity 20.9%; Pred. No. 10; Matches 120; Conservative 79; Mismatches 204; Indels 170; Gaps 28;
QY	3	NYPTVAAQPTTANPLLRHQS	----	FORRELPK-IVETESTSMDITIGSQKQPFLEK 57
DB	805	SYQSATHEQTKAYEEQLAQKLLDLETERILLTKQVAEVAQKDV	-----	852
QY	58	SIDELAAFSVAVETFKROFD	----	DLQKHIEINAIDSKLESNG-----VVLARNN 106
DB	853	CTELDAHKIQODLMQOEKQNSEMEQVKSLTQVYESKLEDGKEQEQTKQILVEKEN 911	----	
QY	107	NHPQMLSPPRNNVSVET	----	ETGNKPEGRMCEL-MCSKGLRKYI 161
DB	912	MILQ-MREGGKKIEILTQKSAKEDSIHILNEEYETKFNQEKMEKVKQKAKEMOETL 970	-----	
QY	162	YANISQO-AKLMEIEP-SALKLAKGPAKEVLCIGKFLYQGRRAFTKESPMSSARQVSL 219	-----	
DB	971	KKLLDQEAALKELENTALELSQKQKQF	-----	NAKMLEAQAANSAGIS 1015
QY	220	ILBSFLMPDRGKGVKIBSWI	----	KDEATAVAWRKELMTTEGLAAAEKN----- 268
DB	1016	DAVSRL	----	ETNQKEQIESLTVHRRRLNDVISWEKLNQ-----AEELQEIHEIQL 1066
QY	269	-----	DARGLLLVACFGVPSNFRSTDLDD	-----LIRMSGSEIAGALKRSQFLVPMVS 318
DB	1067	QEKEQVVAELKQKILLFGCEKEEMKEITWLKEEGVKQDTTLNELOEQKQKS	-----	AHVN 1123
QY	319	GIVBESSIKRGMHTEALEM	-----	VYTFGMEDKFSALVLTSPKMSK 360
DB	1124	SLAQDETCLKAHLKLEVDLNLKSLKENTFLQBLQVLMKLAEDDKRYS-ELTSKLKTTD 1182	-----	
QY	361	ESPERAK	-----	RKAQSLAPK-----BAATKQL----- 384
DB	1183	EEFQSLKSSHERKNSKLEDSLEFKLSBELAQLDICKCKTEALLBEAKTNELINISSK 1242	-----	
QY	385	--AVLSSWQCMETHKLDPAKELPGQIKQIVS-LEKDTLQDK	-----	427
DB	1243	TNALLSRISHCQ--HRTTKVKE--ALLIKTIVSELAQLRLTBEQNTLNSFOQATHQ 1298	-----	
QY	428	MEEKARSLSLM	-----	EEAALAKRMYNQ 451
DB	1299	LEEKENQIKSMKADIESLVTEKEALQKEGNNQ 1331	-----	

RESULT 11  
SNE2 HUMAN  
ID SNE2 HUMAN STANDARD; PRT; 6885 AA.  
AC Q8WXH0; Q8N1S3; Q8NF49; Q8TER7; Q8WNW3; Q8WNW4; Q8WNW5; Q8WXH1;  
DC Q9NU50; Q9UF04; Q9Y2L4; Q9Y4R1;  
PT 15-SEP-2003 (Rel. 42, Created)  
DT 15-SEP-2003 (Rel. 42, Last sequence update)

DT 15-SEP-2003 (rel. 42, Last annotation update)  
 DE Nesprin 2 (Nuclear envelope spectrin repeat protein 2) (Syn-2)  
 DE (Synaptic nuclear envelope protein 2) (Nucleus and actin connecting  
 DE element protein) (NUANCE protein).  
 GN SYNE2 OR NUA OR KIAA1011.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 9), FUNCTION, CHARACTERIZATION, AND  
 RP INTERACTION WITH F-ACTIN.  
 RX MEDLINE=22113122; PubMed=12118075;  
 RA Zhen Y.-Y., Libotte T., Munch M., Noegel A.A., Korenbaum E.;  
 RT "NUANCE, a giant protein connecting the nucleus and actin  
 RT cytoskeleton."; *J. Cell Sci.* 115:3207-3222(2002).  
 RL [2]  
 RN  
 RP SEQUENCE FROM N.A. (ISOFORMS 4; 5 AND 7).  
 RX MEDLINE=21652858; PubMed=11792814;  
 RA Zhang Q., Skepper J.N., Yang F., Davies J.D., Hegyi L., Roberts R.G.,  
 RA Weissberg P.L., Ellis J.A., Shanahan C.M.;  
 RT "Nesprins: a novel family of spectrin-repeat-containing proteins that  
 RT localize to the nuclear membrane in multiple tissues."; *J. Cell Sci.* 114:4485-4498(2001).  
 RL [3]  
 RN  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=22296983; PubMed=12408964;  
 RA Zhang Q., Ragnauth C., Greener M.J., Shanahan C.M., Roberts R.G.;  
 RT "The nesprins are giant actin-binding proteins, orthologous to  
 RT Drosophila melanogaster muscle MSP-300."; *Genomics* 80:473-481(2002).  
 RL [4]  
 RN  
 RP SEQUENCE FROM N.A. (ISOFORM 6).  
 RC TISSUE=Testis;  
 RA Poustka A., Klein M., Mewes H.-W., Gassenhuber J., Wiemann S.;  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX PubMed=12508121;  
 RA Heilig R., Eckenberg R., Petit J.-L., Fonknechten N., Da Silva C.,  
 RA Cattolico L., Levy M., Barbe V., De Berardinis V., Ureta-Vidal A.,  
 RA Pelletier E., Vico V., Anthouard V., Rowen L., Madan A., Qin S.,  
 RA Sun H., Du H., Pepin K., Artiguenave F., Robert C., Craud C.,  
 RA Bruehl T., Jaillon O., Friedlander L., Samson G., Brottier P.,  
 RA Cure S., Segurens B., Aniere F., Samain S., Crepeau H., Abbasi N.,  
 RA Alach N., Boscus D., Dickhoff R., Dors M., Dubois I., Friedman C.,  
 RA Gouyvenoux M., James R., Madan A., Maltrey-Estrada B., Mangenot S.,  
 RA Martins N., Menard M., Oztas S., Ratcliffe A., Shaffer T., Traek B.,  
 RA Vacherie B., Bellemere C., Belser C., Besnard-Gonnet M.,  
 RA Bartol-Mavel D., Boutard M., Briez-Silla S., Combette S.,  
 RA Dufosse-Laurent V., Ferron C., Lechaplais C., Louesse C., Musset D.,  
 RA Magdelenat G., Pateau E., Petit E., Sirvain-Trukniewicz P., Trybou A.,  
 RA Vega-Czarny N., Bataille E., Bluet E., Bordenais I., Dubois M.,  
 RA Dumont C., Guerin T., Hafray S., Hammadi R., Munga J., Pellouin V.,  
 RA Robert D., Wunderle E., Gauguier G., Roy A., Sainte-Marthe L.,  
 RA Verdier J., Verdier-Discala C., Hillier L., Fulton L., McPherson J.,  
 RA Matsuda F., Wilson R., Scarpetti C., Gyapay G., Wincker P., Saurin W.,  
 RA Quetier F., Waterston R., Hood L., Weissbach J.;  
 RT "The DNA sequence and analysis of human chromosome 14."; *Nature* 421:601-607(2003).  
 RL [6]  
 RN  
 RP SEQUENCE FROM N.A. (ISOFORM 8).  
 RC TISSUE=Brain;  
 RX MEDLINE=22338257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.-I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.E., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Uedin T.B., Tohiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.X., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahay J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smallus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences."; *Proc. Natl. Acad. Sci. U.S.A.* 99:16899-16903(2002).  
 RL [7]  
 RN  
 RP SEQUENCE FROM N.A. (ISOFORM 3), AND SEQUENCE OF 1-956 AND 5133-6885  
 RP FROM N.A.  
 RC TISSUE=Spleen, and Tongue;  
 RA Jikuya H., Takano J., Nomura N., Kikuno R., Nagase T., Ohara O.,  
 RA Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,  
 RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,  
 RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,  
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,  
 RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,  
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,  
 RA Kanehori K., Takahashi-Fujii A., Oshima K., Sugiyama A., Kawakami B.,  
 RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;  
 RT "NEO human cDNA sequencing project."; Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 RL [8]  
 RN  
 RP SEQUENCE OF 5709-6885 FROM N.A. (ISOFORM 2).  
 RC TISSUE=Brain;  
 RX MEDLINE=99246063; PubMed=10231032;  
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hitosawa M.,  
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XIII.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro."; *DNA Res.* 6:63-70(1999).  
 RL [9]  
 RN  
 RP REVISIONS.  
 RC TISSUE=Brain;  
 RX MEDLINE=22158633; PubMed=12168954;  
 RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;  
 RT "Construction of expression-ready cDNA clones for KIAA genes: manual  
 RT curation of 330 KIAA cDNA clones."; *DNA Res.* 9:99-106(2002).  
 RL [10]  
 RN  
 RP SEQUENCE OF 5754-6885 FROM N.A.  
 RX MEDLINE=21154917; PubMed=11230166;  
 RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,  
 RA Ansorge W., Boecker M., Bloeker H., Bauersachs S., Blum H.,  
 RA Lauber J., Duesterhoeft A., Beyer A., Koehler K., Strack N.,  
 RA Mewes H.-W., Othenwaelder B., Obermaier B., Tampe J., Heubner D.,  
 RA Wambutt R., Korn B., Klein M., Poustka A.;  
 RT "Towards a catalog of human genes and proteins: sequencing and  
 RT analysis of 500 novel complete protein coding human cDNAs."; *Genome Res.* 11:422-435(2001).  
 RL  
 CC -1- FUNCTION: Involved in the maintenance of nuclear organization and  
 CC structural integrity. Probable anchoring protein which tethers the  
 CC nucleus to the cytoskeleton. Connects nuclei to the cytoskeleton  
 CC by interacting with the nuclear envelope and with F-actin in the  
 CC cytoplasm.  
 CC  
 CC -1- SUBUNIT: Interacts with F-actin via its N-terminal domain.  
 CC  
 CC -1- SUBCELLULAR LOCATION: Type IV membrane protein (Potential). The  
 CC largest part of the protein is cytoplasmic, while its C-terminal  
 CC part is associated with the nuclear envelope, most probably the  
 CC outer nuclear membrane. Remains associated with the nuclear  
 CC envelope during its breakdown in mitotic cells.  
 CC  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event-Alternative splicing; Named isoforms=9;  
 CC Name=1;  
 CC IsoId=Q8WXH0-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q8WXH0-2; Sequence=VSP\_007164, VSP\_007166;  
 CC

CC Notes=No experimental confirmation available;  
 CC Name=3;  
 CC IsoId=Q8WXH0-3; Sequence=VSP\_007155;  
 CC Note=Produced by exon skipping that results in a frameshift. No  
 CC experimental confirmation available;  
 CC Name=4; Synonym=Beta;  
 CC IsoId=Q8WXH0-4; Sequence=VSP\_007156;  
 CC Name=5; Synonym=Alpha;  
 CC IsoId=Q8WXH0-5; Sequence=VSP\_007157, VSP\_007164, VSP\_007165;  
 CC Name=6;  
 CC IsoId=Q8WXH0-6; Sequence=VSP\_007158, VSP\_007165, VSP\_007166;  
 CC Note=No experimental confirmation available;  
 CC Name=7; Synonym=Gamma;  
 CC IsoId=Q8WXH0-7; Sequence=VSP\_007154, VSP\_007163;  
 CC Name=8;  
 CC IsoId=Q8WXH0-8; Sequence=VSP\_007161, VSP\_007162;  
 CC Note=No experimental confirmation available;  
 CC Name=9; Synonym=Nuance-N-33;  
 CC IsoId=Q8WXH0-9; Sequence=VSP\_007159, VSP\_007160;  
 CC - TISSUE SPECIFICITY: Widely expressed with higher level in kidney,  
 CC adult and fetal liver, stomach and placenta. Weakly expressed in  
 CC skeletal muscle and brain. Isoform 5 is highly expressed in  
 CC pancreas, skeletal muscle and heart.  
 CC - DOMAIN: The Klarsicht domain mediates the nuclear envelope  
 CC targeting.  
 CC - SIMILARITY: Belongs to the Nesprin family.  
 CC - SIMILARITY: Contains 1 actin-binding domain.  
 CC - SIMILARITY: Contains 2 calponin-homology (CH) domains.  
 CC - SIMILARITY: Contains 1 Klarsicht domain.  
 CC - SIMILARITY: Contains 13 leucine-rich (LRR) repeats.  
 CC - SIMILARITY: Contains 9 spectrin repeats.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC  
 CC EMBL; AF435010; AAL33547.1; -  
 CC EMBL; AF435011; AAL33548.1; -  
 CC EMBL; AY061757; AAL33800.1; -  
 CC EMBL; AY061758; AAL33801.1; -  
 CC EMBL; AY061759; AAL33802.1; -  
 CC EMBL; AF495911; AANG0443.1; -  
 CC EMBL; AL117404; CAB55903.1; -  
 CC EMBL; AL162832; -; NOT\_ANNOTATED\_CDS.  
 CC EMBL; AL355094; -; NOT\_ANNOTATED\_CDS.  
 CC  
 CC Query Match 4.0%; Score 125.5; DB 1; Length 6885;  
 CC Best Local Similarity 18.9%; Pred. No. 50;  
 CC Matches 100; Conservative 89; Mismatches 190; Indels 149; Gaps 20;  
 CC  
 CC 57 KSIDLAAPSVAVETFKQFDLO-----KHISIENAIKSL 94  
 CC 3214 KORENSEADVETKREFDELQVQLNTSILRTNVLDAYENLRYKAVRAVESIT 3273  
 CC  
 CC 95 ESNQGVLAARN--NFHQPMLSPRRNVSVETVY-----VSQPS----- 132  
 CC 3274 SLEAIIRYDVGNPEBSLEMLPKQBELESTVAHQDLTEKLGMTSSPEAKLQYTL 3333  
 CC  
 CC 133 QIVPETSNGEGGCMELMCKGLRY-----IYANISDQAKLMEIIPSAKLAK 183  
 CC 3334 QELVSKNGAMKEAFQAQTEAEKRYLENKYCKRMKEEDINYLSKMETVIGQSSSPLSY 3393  
 CC  
 CC 184 EPAKFLVDCIGKFLQGRRAFTKSPSSRAQVSLLESFLMPDRGKGKVI-----E 238  
 CC 3394 REALERLE--QSKALVSNLSITKBEELMKRLQILRLR---CTENDGICLLKIVSALWE 3448  
 CC  
 CC 239 SWIK--DEAETAVARKELMTGEGGLAAEKVDARGLILLVACGVSPNFS----- 288  
 CC 3449 KWSLSLEAKENWMECEKLGKWF--VSEIEIERAILDNLQBELPEISKTEAATTEL 3507

## RESULT 12

## BEM3 YEAST

ID BEM3\_YEAST STANDARD; PRT: 1128 AA.

AC P32873;

DT 01-OCT-1993 (Rel. 27, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE GTPase-activating protein BEM3.

GN BEM3 OR YPL115C OR LPH12C.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI\_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=94043316; PubMed=8227021;

RA Zheng Y., Hart M.J., Shinjo K., Evans T., Bender A., Cerione R.A.;

RT "Biochemical comparisons of the Saccharomyces cerevisiae Bem2 and

RT Bem3 proteins. Delineation of a limit Cdc42 GTPase-activating protein

RT domain.";

RL J. Biol. Chem. 268:24629-24634(1993).

RN [2]

RP SEQUENCE FROM N.A.

RX STRAIN=S28C / AB972;

RX MEDLINE=97313271; PubMed=9169875;

RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansoorge W.,

RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,

RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,

RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,

RA Dietrich F.S., Delius H., Dipaolo T., Dubois E., Duesterhoeft A.,

RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,

RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.,

RA Hunnicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,

RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,

RA Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,

RA Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,

RA Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,

RA Scherens B., Schramm S., Schroeder M., Sidcu A.M., Tettelin H.,

RA Urrutazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,

RA Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,

RA Zhong W.W., Zollner A., Vo D.H., Hani J.;

RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";

RL Nature 387:103-105(1997).

CC - FUNCTION: GTPASE-ACTIVATING PROTEIN (GAP) FOR CDC42 AND LESS

CC EFFICIENTLY FOR RHO1 NEGATIVE REGULATOR OF THE PHEROMONE-RESPONSE

CC PATHWAY THROUGH THE STE20 PROTEIN KINASE.

CC - SUBCELLULAR LOCATION: Cytoplasmic.

CC - SIMILARITY: TO REGULATORY DOMAIN OF PROTEIN KINASE C AND PCR,

CC THE PRODUCT OF THE BREAKPOINT CLUSTER REGION GENE.

CC - SIMILARITY: Contains 1 PH domain.

CC - SIMILARITY: Contains 1 Rho-GAP domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC ENBL; L14558; AAB34453.1; --  
 DR ENBL; U43503; AAB68247.1; --  
 DR PIR; A49960; A49960.  
 DR SGD; S0006036; BEM3.  
 DR GO; GO:0005100; F:rho GTPase activator activity; IDA.  
 DR GO; GO:0004871; F:signal transducer activity; IPI.  
 DR GO; GO:0007118; P:apical bud growth; IPI.  
 DR GO; GO:0000283; P:establishment of cell polarity (sensu Sacch. .); IPI.  
 DR GO; GO:0007125; P:invasive growth; IPI.  
 DR GO; GO:0007119; P:isotropic bud growth; IPI.  
 DR GO; GO:0007124; P:pseudohyphal growth; IPI.  
 DR GO; GO:0007264; P:small GTPase mediated signal transduction; IPI.  
 DR InterPro; IPR001849; PH.  
 DR InterPro; IPR001683; PX.  
 DR InterPro; IPR000198; RhogAP.  
 DR Pfam; PF00169; PH; 1.  
 DR Pfam; PF00787; PX; 1.  
 DR Pfam; PF00620; RhogAP; 1.  
 DR SMART; SM00233; PH; 1.  
 DR SMART; SM00312; PX; 1.  
 DR SMART; SM00324; RhogAP; 1.  
 DR PROSITE; PS50003; PH DOMAIN; 1.  
 DR PROSITE; PS50238; RhogAP; 1.  
 KW GTPase activation.  
 FT DOMAIN 634 741 PH.  
 FT DOMAIN 913 1128 RHOGAP.  
 SQ SEQUENCE 1128 AA; 124912 MW; 4BF03EAD6EF10283 CRC64;

Query Match 4.0%; Score 125; DB 1; Length 1128;  
 Best Local Similarity 19.3%; Pred. No. 4.7; Indels 220; Gaps 30;  
 Matches 134; Conservative 97; Mismatches 243;

QY 13 TTTANLLQHQSEQRRLPKIVITESTSDITIGSKQFQF-----LKSIDEAAAF 65  
 DB 12 STTLELLAQVNDHRSKKDKSIEHIEKGTCS-----GKRNFSYDEIFTENIKLVQVEY 66  
 QY 66 SVAVETFKQFDLDQV-----IESIEN-AIDSKLESNGVLAARN-----N 107  
 DB 67 ETEIESLEKVIDMLQKREASLEVLEQVQNDSDRSYVNDQSFVLPFPAERKAHISLN 126  
 QY 108 FHOPLMLSP-----RNVSVTITVTSQFSGEIVPETS-NKPEGGRM-----CELMCSKGLR 158  
 DB 127 LPIPTLSPLQOGSDVALETSVTPVQIGVTSNTSISRKLQNLNMLNDEIANSFSSP 186  
 QY 159 KTIYANISDQAKMEF-----IPSAKLAKPAKAVL----- 190  
 DB 187 KIINRSVSPTKIHSQLASPAASVYTTTSTRITIKSPNKGSKPLQERLRSFQPNRMTA 246  
 QY 191 -----DCICK-----FVLOGRAFTKESPMSSARQ 215  
 DB 247 VINNHLSPLKASTNNLDLTSKSOQLTNDALQKNDRVSSITSSAYTTGTTSRAKS 306  
 QY 216 VSLILEFLMPDRGKGKVKIESWIKDEAETAAVAMVKRLMTEGGLAAAEKMDARGLL 275  
 DB 307 PSSLL-----EVK-----EGENKALGFSP-----ASKELDDFTCLL 338  
 QY 276 LVACFGVPNSRSTDLDL-----IRMSGNETAGALKRSQFLVPMVSGIVESSIKRG 328  
 DB 339 -----DSFGBEDLVNTSDKPLSIKSTINESUPPPAPTFFSPPTSGNINKNSTPLS 391  
 QY 329 MHTALEMYTTFQMEDKFSAL-----VLTSFLKMSKESFERAKRKAQ---SPLAFKE 378  
 DB 392 SHLAS--PVLNKKDNDFGAQAQAKLKKRFLVLTSSLPNLSTKLSTTSQNASLPPNPPVESS 449  
 QY 379 AATKQALAVLSVMQCNETHKLPAPKELPGWQIQEIVSLEKDTLQDKEMEKAARSLSLM 438  
 DB 450 SKOKQLGETASI---HSTNTLNTFTSTPGQSLK-----TLR-----RPHASSVSTV 492

QY 439 EBAALAKMYNQIKRP---RLSPMEPPVTSSSYPIYDRSPFSPQDDDDQDEISALVS 495  
 DB 493 KQVA-----QSLKSDIFLFWQPEDFTQIQIEVLSTLYRD-----NEDDLSLIAIID 539  
 QY 496 SYLGPS-----TSFPHRSRSPEYM-----VPLPHGGLGRSVVAYEHLAFNSYSPGHGR 545  
 DB 540 RKSQKEMPKFKSKSHKVKRELDVYKSHVDFPLP-----TLPD-----RQ 579  
 QY 546 LHRQYSPSLVHGQRHPL-QYSPPIHGQQQLPYGI 578  
 DB 580 LFQTLSPKVDTRKNILNQYVTSIFSVPFKNV 613

RESULT 13  
 CING\_HUMAN  
 ID CING\_HUMAN STANDARD; PRT; 1197 AA.  
 AC Q9P2M7; Q9NR25;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cingulin.  
 GN CCN OR KIAA1319.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]\_TaxID=9606;  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Neuroepithelium;  
 RX MEDLINE=20499514; PubMed=1042084;  
 RA Citi S., D'Atri F., Parry D.A.D.;  
 RT "Human and Xenopus cingulin share a modular organization of the  
 RT coiled-coil rod domain: predictions for intra- and intermolecular  
 RT assembly.";  
 RL J. Struct. Biol. 131:135-145 (2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=20181126; PubMed=10718198;  
 RA Nagase T., Kikuno R., Ishikawa K.-I., Hirosewa M., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XVI.  
 RT The complete sequences of 150 new cDNA clones from brain which code  
 RT for large proteins in vitro.";  
 RL DNA Res. 7:165-73 (2000).  
 RN [3]  
 RP INTERACTION WITH TJPI.  
 RX MEDLINE=22140336; PubMed=12023291;  
 RA D'Atri F., Nadalutti F., Citi S.;  
 RT "Evidence for a functional interaction between cingulin and ZO-1 in  
 RT cultured cells.";  
 RL J. Biol. Chem. 277:27757-27764 (2002).  
 CC -!- FUNCTION: Probably plays a role in the formation and regulation of  
 CC the tight junction (TJ) paracellular permeability barrier.  
 CC -!- SUBUNIT: Homodimer (By similarity). Interacts with TJPI/ZO-1.  
 CC -!- TISSUE SPECIFICITY: Localized on the cytoplasmic face of tight  
 CC junctions of polarized epithelia and some endothelia. Expressed in  
 CC pancreas, kidney, liver and lung, but not in skeletal muscle,  
 CC placenta, brain or heart.  
 CC -!- DOMAIN: Deletion of the ZO-1 interaction motif (ZIM) decreases but  
 CC does not abolish colocalization with ZO-1.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC ENBL; AF263462; AAF74498.1; ALT\_INIT.  
 CC ENBL; AB037740; BAA92557.1; ALT\_INIT.  
 DR InterPro; IPR002928; Myosin\_tail.  
 DR Pfam; PF01576; Myosin\_tail; 1.

KW Tight junction; Coiled coil. HEAD.  
FT DOMAIN 1 351  
FT DOMAIN 352 1154  
FT DOMAIN 1155 1197  
FT DOMAIN 363 836  
FT SITE 42 56  
FT SITE 106 400  
FT SITE 1197 AA; 136385 MW; OC9375283ABAAF3D CRC64;  
SEQUENCE 1197 AA; 136385 MW; OC9375283ABAAF3D CRC64;  
Query Match 4.0%; Score 124.5; DB 1; Length 1197;  
Best Local Similarity 19.9%; Pred. No. 5.5; Indels 143; Gaps 25;  
Matches 111; Conservative 91; Mismatches 213;  
QY 7 TVAAOPTTANPLLRHOS-----EORRR-----ELPKIV----- 36  
Db 163 TIDTAPLSVDSLINKFDSOLGGQARGTGRTRMLPEQKRKSKSLDRLPTDPEERE 222  
QY 37 -----ETESTMDITICGKQPLKSIDELAFSAVAVTFKQPD--LQKHISIE 88  
Db 223 ROSTNHTSSTKYDHHVGTGSKQPAQSQNSLPSGFSRS-----RQTDWVLOSFEFPRS 277  
QY 89 AID-SKLENGVVLAAARNNHOPMLSPRNVSVEITV--TVSQPSQEIIVPETSNNKPEG 145  
Db 278 AQPPTWLFKSTPDLRLD---QEEAAPGSDVDMKATIYGLREGSSE--SETSVRKV 331  
QY 146 GRMCE-----LMCKGLKRYIVA-----NTSD-QAKLMEI-----PSALKLAKEPA 186  
Db 332 SILVKEMOPLVMVSSGSTKAVAGQGLTRKVEELQRLDEEVKKRQKLEPFOVGLERQLE 391  
QY 187 KVLDCIGKFKYLOGRATTKESPMSSARQVSLILLESFLMPDRKPKGKVKIESWIKDEAE 246  
Db 392 EKTBCSLQLELLER---KGEAQSNKE-----LQNKRLLDQGED-----LRHGLE 436  
QY 247 TAAVAWRKELMTTEGLAAEKMDARGLLLVACFGVPSNFRSTDLDLIRMSGNEIAGA 306  
Db 437 TQWMLQNKLVKHVQPEPAKEVLLKDL-----ETRELLBEV-LECKQVVEEQ 483  
QY 307 LK-RSQFLVPTVSGIVSSIKRGHIEALEVMYTFGMEKFSALVLTSPFKMSKESFER 365  
Db 484 LRERRELTALKALKEEVASRDQVEHVROQY-----QRDTEQLRR 525  
QY 366 AKRKAQSPFAPKEAATKQALVLSVMQCMETHKLDPAKELFQ-WQ-----IKE 412  
Db 526 SMQDATQDHAVLEAERQMSALVRGLQ---RELSETSEETHQWQFMQKXEDLRATQ 581  
QY 413 QIVSLEKDTQLDQKEMEKARSLMEEAALAKRMVNOI---KRPILSPME-----MPP 464  
Db 582 ELQLRMKEEMEEELGKIEVLORELFQARASAGDTQVVEVLKKEILRTQELKELQAE 641  
QY 465 VTSSSYSPYDRSFPFSQ 482  
Db 642 RSQEVAGRHRRDRELEKQ 659  
RESULT 14  
PCLO CHICK STANDARD; PRT; 5120 AA.  
AC Q9PU36;  
DT 28-FEB-2003 (rel. 41, Created)  
DT 28-FEB-2003 (rel. 41, Last sequence update)  
DT 28-FEB-2003 (rel. 41, Last annotation update)  
DE Ficollo protein (Aczonin) (Fragment).  
GN PCLO OR ACZ.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=99439764; PubMed=10508862;  
RA Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,

RA Killmann M.W.;  
RT "Aczonin, a 550-kd putative scaffolding protein of presynaptic active  
RT zones, shares homology regions with rim and bassoon and binds  
RT profilin";  
RL J. Cell Biol. 147:151-162(1999).  
CC -!- FUNCTION: May act as a scaffolding protein involved in the  
CC organization of synaptic active zones and in synaptic vesicle  
CC trafficking (By similarity).  
CC -!- SUBUNIT: Interacts with Rabac1/Prai and profilin (By similarity).  
CC -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of  
CC synaptic junctions (By similarity).  
CC -!- DOMAIN: C2 domain 1 is involved in binding calcium and  
CC phospholipids. Calcium binds with low affinity but with high  
CC specificity and induces a large conformational change.  
CC -!- SIMILARITY: Contains 2 C2 domains.  
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; Y19187; CAB60725.1; --  
CC HSSP; P04410; I425.  
CC GO; GO:0045202; C:synaptic junction; ISS.  
CC GO; GO:0005509; F:calcium ion binding activity; ISS.  
CC GO; GO:0005544; F:calcium-dependent phospholipid binding acti...; ISS.  
CC GO; GO:0005522; F:profilin binding activity; ISS.  
CC GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.  
CC GO; GO:0016080; P:synaptic vesicle targeting; ISS.  
CC InterPro; IPR000008; C2.  
CC InterPro; IPR001478; PDZ.  
CC InterPro; IPR001565; Synaptotagmin.  
CC Pfam; PF00168; C2; 2.  
CC PRINTS; PR00399; SYNAPTOTAGMN.  
CC SMART; SM00239; C2; 2.  
CC SMART; SM00238; PDZ; 1.  
CC PROSITE; PS00499; C2\_DOMAIN\_1; 1.  
CC PROSITE; PS00004; C2\_DOMAIN\_2; 2.  
CC PROSITE; PS0106; PDZ; 1.  
CC Calcium/phospholipid-binding; Zinc; Metal-binding; Zinc-finger;  
KW Repeat.  
FT NON\_TER  
FT DOMAIN 1 357  
FT DOMAIN 258 357  
FT ZN\_FING 368 392  
FT ZN\_FING 836 859  
FT DOMAIN 2324 2343  
FT DOMAIN 4414 4493  
FT DOMAIN 4627 4726  
FT DOMAIN 5003 5094  
FT DOMAIN 5120 AA; 560751 MW; A638D9891B658412 CRC64;  
SEQUENCE 5120 AA; 560751 MW; A638D9891B658412 CRC64;  
Query Match 4.0%; Score 123.5; DB 1; Length 5120;  
Best Local Similarity 17.9%; Pred. No. 45;  
Matches 129; Conservative 115; Mismatches 296; Indels 181; Gaps 27;  
QY 4 YPPTVAAOP-TTANPLLRHOSQRRREL---PKIVETESTMDITICGSK----- 51  
Db 3229 YDPSGTGSPQTDDQALLEGQYATAENGQFWPTDDATTATGAGVIGIEISQSYTWYVQSD 3288  
QY 52 -----QPQFLKSIDELAAFSVAVTFKR-----QFDDLOKHIES-----IEN 88  
Db 3289 GITQYIPRSGILSVSEMSLKDIDVREKQKKRSMPLKRGVVELESLSEEPKCYKK 3348  
QY 89 AIDSKLENGVVLAAARNNHOPMLSPRNVSVEITVVSQPSQ---EIVPETSNNKPEG 145  
Db 3349 IVDVSGVQTDDEGDGRGYTNRR-----RRTKKSVDTSVQTDDEQDEWDLSSRSGRPRV 3403

Genome Res. 11:1053-1070(2001).

[3] SEQUENCE FROM N.A. (ISOFORM 3).  
TTSUVE=phochromytoma;  
Li Y., Peng Y., Li N., Gu W., Han Z., Fu G., Chen Z.;  
"A novel gene expressed in human pheochromocytoma."  
Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

[4] SEQUENCE FROM N.A.  
MEDLINE=20057165; PubMed=10591208;  
Dunham I., Hunt A.R., Collins J.E.,  
Clamp M., Smink L.J., Alnsough R., Almeida J.P., Babbage A.,  
Bagnoley C., Bailey J., Barlow K., Bates K.N., Beasley O., Bird C.P.,  
Blakey S., Bridgman A.M., Buck D., Burgess J., Burrill W.D.,  
Burton J., Carder C., Carter N.P., Chen Y., Clark G., Clegg S.M.,  
Coble V., Cole C.G., Collier R.E., Connor R.E., Conroy D., Corby N.,  
Coville G.J., Cox A.V., Davis J., Dawson E., Dhami P.D., Dockree C.,  
Dodsworth K., Durbin R.M., Ellington A., Evans K.L., Fey J.M.,  
Fleming S., French L., Garner A.A., Gilbert J.G.R., Goward M.E.,  
Graham D., Griffiths M.N., Hall C., Hall R., Hall-Tamlyn J.,  
Heathcott R.W., Ho S., Holmes S., Hunt S.E., Jones M.C., Kershaw J.,  
Kimberley A., King A., Laird G.K., Langford C.P., Leversha M.A.,  
Lloyd C., Lloyd D.M., Martyn I.D., Mashreghi-Mohammadi M.,  
Matthews L., McCann O.T., Mcclay J., McLaren S., Mcmurray A.A.,  
Milne S.A., Mortimore B.J., Odell C.N., Pavitt R., Pearce A.V.,  
Pearson D., Phillips B.J., Phillips S.H., Plumb R.W., Ramsay H.,  
Ramsey Y., Rogers L., Ross M.T., Scott C.E., Sehra H.K., Stuce C.D.,  
Snalley S., Smith M.L., Soderlund C., Spragon L., Steward C.A.,  
Suiston J.E., Swann R.M., Vaudin M., Wall M., Wallis J.M.,  
Whiteley M.N., Willey D., Williams L., Williams S., Williamson H.,  
Wilmer T.E., Wilming L., Wright C.L., Rubbard T., Bentley D.R.,  
Beck S., Rogers J., Shimizu N., Minoshima S., Kawasaki K., Sasaki T.,  
Aekawa S., Kudoh J., Shintani A., Shibuya K., Yoshizaki Y., Aoki N.,  
Mitsuyama S., Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S.,  
Do A., Do T., Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S.,  
Lai H., Lao H.L., Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E.,  
Nguyen T., Pan H., Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S.,  
Sloan D., Song L., Wang Q., Wang Y., Wang Z., White J., Willingham D.,  
Wu H., Yao Z., Zhan M., Zhang G., Chissoe S., Murray J., Miller N.,  
Minx P., Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H.,  
Bourne S., Cordes M., Du Z., Fulton L., Goela D., Graves T.,  
Hawkins J., Hinds K., Kemp K., Latreille P., Layman D., Ozersky P.,  
Rohlfing T., Schect P., Walker C., Wamstley A., Wohlmann P., Pepin K.,  
Nelson J., Korfei I., Bedell J.A., Hillier L., Mardis E., Waterston R.,  
Wilson R., Emanuel S., Shaikh T., Kurahashi H., Saitta S.,  
Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,  
Edgarman L., Kim U.J., Shizuya H., Simon M.I., Dumaneki J.P.,  
Peyraud M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,  
O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,  
Khan A.S., Lane L., Tiliakou Y., Wright H.;  
"The DNA sequence of human chromosome 22."  
Nature 402:489-495(1999).

[5] SEQUENCE FROM N.A. (ISOFORM 2).  
TISSUE=Eye, and Skin;  
MEDLINE=22388257; PubMed=12477932;  
Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
Altschul S.F., Zseberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
Hopkins R.F., Jansz S., Klotzel T., Max A.I., Wang J., Hsieh F.,  
Diachenko L., Marushina K., Farmer A.A., Rubin G.M., Hong L.,  
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
Villalon D.K., Munz D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
Blakesley R.W., Tuchman J.W., Green E.D., Dickson M.C.,  
Rodriguez A.C., Grumwood J., Schmutz J., Myers R.M.,  
Butterfield Y.S.N., Krzywicki M.I., Skalska U., Smalls D.E.,  
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 14, 2003, 10:34:34 ; Search time 107 seconds  
(without alignments)  
1468.729 Million cell updates/sec

Title: US-09-890-475-1  
Perfect score: 3104  
Sequence: 1 MSNYPPTVAQPTTTANPLL.....RYLGLSNQSPRNSSLDPK 609

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL.23.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_virus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	931	30.0	196	Q8RV21	arabidopsis
2	927	29.9	186	Q8RUS4	arabidopsis
3	921	29.7	186	Q8RU84	arabidopsis
4	921	29.7	186	Q8RUZ8	arabidopsis
5	921	29.7	186	Q8RW66	arabidopsis
6	697	22.5	138	Q8RVCO	arabidopsis
7	574	18.5	104	Q8RV51	arabidopsis
8	340	11.0	558	Q8L7D6	arabidopsis
9	296.5	9.6	532	Q8LUV4	arabidopsis
10	295	9.5	473	Q8C6S2	arabidopsis
11	286.5	9.2	862	Q84649	arabidopsis
12	282	9.1	532	Q840H8	arabidopsis
13	277	8.9	505	Q8W4S7	arabidopsis
14	262.5	8.5	1337	Q8LV73	arabidopsis
15	245	7.9	470	Q8FFL1	arabidopsis
16	234.5	7.6	507	Q83345	arabidopsis

17	156.5	5.0	3117	4	Q8WY20	homo sapien
18	151.5	4.9	2594	5	Q9VM88	drosohila
19	150	4.8	1088	4	Q8IUD3	homo sapien
20	149.5	4.8	948	11	Q8CIY9	rattus norv
21	146	4.7	992	4	Q8IUD5	homo sapien
22	146	4.7	1003	4	Q9UPS1	homo sapien
23	145.5	4.7	948	4	Q9UIK7	homo sapien
24	143.5	4.6	1116	4	Q8IUD2	homo sapien
25	141.5	4.6	1248	3	Q8XOV7	neurospora
26	140.5	4.5	775	4	Q9ECN9	homo sapien
27	139.5	4.5	1201	5	O16844	drosohila
28	138	4.4	778	11	Q8CCR3	mus musculus
29	138	4.4	1291	17	Q8T8F2	pyrococcus
30	137.5	4.4	2074	11	Q88542	mus musculus
31	137	4.4	809	11	Q9D4H2	mus musculus
32	136.5	4.4	712	5	Q8IH04	drosohila
33	136.5	4.4	1201	5	Q9V4K0	drosohila
34	135	4.3	976	11	Q99MI2	mus musculus
35	135	4.3	1120	11	Q99MI1	mus musculus
36	134	4.3	841	10	Q23037	arabidopsis
37	132.5	4.3	1762	10	Q94DC2	oryza sativ
38	132	4.3	1316	16	Q8D2C0	wiggleswort
39	132	4.3	5327	5	O76891	drosohila
40	131.5	4.2	993	4	Q9NWP9	homo sapien
41	131.5	4.2	1249	4	Q9HCV1	homo sapien
42	130.5	4.2	981	4	O15014	homo sapien
43	130.5	4.2	5412	5	Q9W596	drosohila
44	128	4.1	2238	11	O70365	mus musculus
45	127.5	4.1	1005	4	Q9P216	homo sapien

ALIGNMENTS

RESULT 1

Q8RV21	PRELIMINARY;	PRT;	186 AA.
AC	Q8RV21;		
DT	01-JUN-2002 (TREMELrel. 21, Created)		
DT	01-JUN-2002 (TREMELrel. 21, Last sequence update)		
DT	01-JUN-2002 (TREMELrel. 21, Last annotation update)		
DE	FRIGIDA protein (Fragment).		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	surceids II; Brassicales; Brassicaceae; Arabidopsis.		
OX	NCBI_TaxID=3702;		
RN	[1]		

RP	SEQUENCE FROM N.A.		
RC	STRAIN=cv. RSCH-4, cv. GOT-32, cv. TSU-0, and cv. Ler;		
RA	Hagenblad J., Nordborg M.;		
RT	"Sequence Variation and Haplotype Structure Surrounding the Flowering		
RT	Time Locus FRI in Arabidopsis thaliana."		
RL	Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AY092681; AAM12134.1; -		
DR	EMBL; AY092683; AAM12136.1; -		
DR	EMBL; AY092691; AAM12144.1; -		
DR	EMBL; AY092693; AAM12146.1; -		
FT	NON_TER	1	
FT	NON_TER	186	
SQ	SEQUENCE	186 AA; 20727 MW; 491F6F2C7AB3229B CRC64;	

Query Match 30.0%; Score 931; DB 10; Length 186;  
Best Local Similarity 99.5%; Pred. No. 1.1e-57;  
Matches 185; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy	34	KIVETESTSMDITIGSQKOPFLKSIDBLAAAFSAVETFKQFDLQKHIESIENAI	DSK 93
Db	1	KIVETESTSMDITIGSQKOPFLKSIDBLAAAFSAVETFKQFDLQKHIESIENAI	DSK 60
Qy	94	LESNGVVLAAARNNNFHQPMLSPPRNNVSVETTVTSQPSQBIIVPETS	NKPEGGRMCLMC 153
Db	61	LESNGVVLAAARNNNFHQPMLSPPRNNVSVETTVTSQPSQBIIVPETS	NKPEGGRMCLMC 120

```

QY 154 SKGLRKYIYANISDQAKLMEIIPSAKLAKEPAKFVLDICIGKFLYQGRRAFTKESPMSSA 213
DB 121 SKGLRKYIYANISDQAKLMEIIPSAKLAKEPAKFVLDICIGKFLYQGRRAFTKESPMSSA 180
QY 214 RQVSL 219
DB 181 RQVSL 186

RESULT 2
Q8RUS4
ID Q8RUS4 PRELIMINARY; PRT; 186 AA.
AC Q8RUS4;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DE FRIGIDA protein (Fragment).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;

[1]
SEQUENCE FROM N.A.
RP STRAIN-cv. MT-0, cv. Koln, cv. Col, cv. NC-6, and cv. DEM-4;
RC Hagenblad J., Nordborg M.;
RT "Sequence Variation and Haplotype Structure Surrounding the Flowering
Time Locus FRI in Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY092678; AAM12131.1; -
DR EMBL; AY092680; AAM12133.1; -
DR EMBL; AY092688; AAM12141.1; -
DR EMBL; AY092689; AAM12142.1; -
DR EMBL; AY092690; AAM12143.1; -
FT NON_TER 1
FT NON_TER 186
SQ SEQUENCE 186 AA; 20781 MW; F809E401DEDC1CDC CRC64;

Query Match 29.7%; Score 921; DB 10; Length 186;
Best Local Similarity 98.9%; Pred. No. 5.3e-57;
Matches 184; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 34 KIVETESTSMITIGQSKQPFKLSIDELAFAFVAVETFKRQFDDLKQHIESIENAI DSK 93
DB 1 KIVETESTSMITIGQSKQPFKLSIDELAFAFVAVETFKRQFDDLKQHIESIENAI DSK 60

[1]
SEQUENCE FROM N.A.
RP STRAIN-Various strains;
RC Hagenblad J., Nordborg M.;
RT "Sequence Variation and Haplotype Structure Surrounding the Flowering
Time Locus FRI in Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY092677; AAM12130.1; -
DR EMBL; AY092679; AAM12132.1; -
DR EMBL; AY092694; AAM12147.1; -
DR EMBL; AY092695; AAM12148.1; -
DR EMBL; AY092696; AAM12149.1; -
FT NON_TER 1
FT NON_TER 186
SQ SEQUENCE 186 AA; 20693 MW; 8ADC65A63EA2FC80 CRC64;

Query Match 29.9%; Score 927; DB 10; Length 186;
Best Local Similarity 99.5%; Pred. No. 2e-57;
Matches 185; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 34 KIVETESTSMITIGQSKQPFKLSIDELAFAFVAVETFKRQFDDLKQHIESIENAI DSK 93
DB 1 KIVETESTSMITIGQSKQPFKLSIDELAFAFVAVETFKRQFDDLKQHIESIENAI DSK 60

[1]
SEQUENCE FROM N.A.
RP STRAIN-cv. MT-0, cv. Koln, cv. Col, cv. NC-6, and cv. DEM-4;
RC Hagenblad J., Nordborg M.;
RT "Sequence Variation and Haplotype Structure Surrounding the Flowering
Time Locus FRI in Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY092677; AAM12130.1; -
DR EMBL; AY092679; AAM12132.1; -
DR EMBL; AY092694; AAM12147.1; -
DR EMBL; AY092695; AAM12148.1; -
DR EMBL; AY092696; AAM12149.1; -
FT NON_TER 1
FT NON_TER 186
SQ SEQUENCE 186 AA; 20693 MW; 8ADC65A63EA2FC80 CRC64;

Query Match 29.9%; Score 927; DB 10; Length 186;
Best Local Similarity 99.5%; Pred. No. 2e-57;
Matches 185; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 34 KIVETESTSMITIGQSKQPFKLSIDELAFAFVAVETFKRQFDDLKQHIESIENAI DSK 93
DB 1 KIVETESTSMITIGQSKQPFKLSIDELAFAFVAVETFKRQFDDLKQHIESIENAI DSK 60

[1]
SEQUENCE FROM N.A.
RP STRAIN-cv. MT-0, cv. Koln, cv. Col, cv. NC-6, and cv. DEM-4;
RC Hagenblad J., Nordborg M.;
RT "Sequence Variation and Haplotype Structure Surrounding the Flowering
Time Locus FRI in Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY092678; AAM12131.1; -
DR EMBL; AY092680; AAM12133.1; -
DR EMBL; AY092688; AAM12141.1; -
DR EMBL; AY092689; AAM12142.1; -
DR EMBL; AY092690; AAM12143.1; -
FT NON_TER 1
FT NON_TER 186
SQ SEQUENCE 186 AA; 20688 MW; 591214F7A5E24DAA CRC64;

Query Match 29.7%; Score 921; DB 10; Length 186;
Best Local Similarity 98.9%; Pred. No. 5.3e-57;
Matches 184; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 34 KIVETESTSMITIGQSKQPFKLSIDELAFAFVAVETFKRQFDDLKQHIESIENAI DSK 93
DB 1 KIVETESTSMITIGQSKQPFKLSIDELAFAFVAVETFKRQFDDLKQHIESIENAI DSK 60

```

```

[1]
SEQUENCE FROM N.A.
RP STRAIN-cv. MT-0, cv. Koln, cv. Col, cv. NC-6, and cv. DEM-4;
RA Hagenblad J., Nordborg M.;
RT "Sequence Variation and Haplotype Structure Surrounding the Flowering
Time Locus FRI in Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY092678; AAM12131.1; -
DR EMBL; AY092680; AAM12133.1; -
DR EMBL; AY092688; AAM12141.1; -
DR EMBL; AY092689; AAM12142.1; -
DR EMBL; AY092690; AAM12143.1; -
FT NON_TER 1
FT NON_TER 186
SQ SEQUENCE 186 AA; 20781 MW; F809E401DEDC1CDC CRC64;

Query Match 29.7%; Score 921; DB 10; Length 186;
Best Local Similarity 98.9%; Pred. No. 5.3e-57;
Matches 184; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 34 KIVETESTSMITIGQSKQPFKLSIDELAFAFVAVETFKRQFDDLKQHIESIENAI DSK 93
DB 1 KIVETESTSMITIGQSKQPFKLSIDELAFAFVAVETFKRQFDDLKQHIESIENAI DSK 60

[1]
SEQUENCE FROM N.A.
RP STRAIN-cv. MT-0, cv. Koln, cv. Col, cv. NC-6, and cv. DEM-4;
RC Hagenblad J., Nordborg M.;
RT "Sequence Variation and Haplotype Structure Surrounding the Flowering
Time Locus FRI in Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY092678; AAM12131.1; -
DR EMBL; AY092680; AAM12133.1; -
DR EMBL; AY092688; AAM12141.1; -
DR EMBL; AY092689; AAM12142.1; -
DR EMBL; AY092690; AAM12143.1; -
FT NON_TER 1
FT NON_TER 186
SQ SEQUENCE 186 AA; 20688 MW; 591214F7A5E24DAA CRC64;

Query Match 29.7%; Score 921; DB 10; Length 186;
Best Local Similarity 98.9%; Pred. No. 5.3e-57;
Matches 184; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 34 KIVETESTSMITIGQSKQPFKLSIDELAFAFVAVETFKRQFDDLKQHIESIENAI DSK 93
DB 1 KIVETESTSMITIGQSKQPFKLSIDELAFAFVAVETFKRQFDDLKQHIESIENAI DSK 60

```

```

QY 94  LEANGVVLARNNNFHQPMLSPPNNVSVETTTVVSQPSQEIIVPETSNNKPEGGRMCCLMC 153
|
|
|
Db 61  LEANGVVLARNNNFHQPMLSPPNNVSVETTTVVSQPSQEIIVPETSNNKPEGGRMCCLMC 120
|
|
|
QY 154  SKGLRKYYIANISDQAKLMEIIPSAKLAKPAKFVLDICIGKIFYLQGRRAFTKESPMSSA 213
|
|
|
Db 121  SKGLRKYYIANISDQAKLMEIIPSAKLAKPAKFVLDICIGKIFYLQGRRAFTKESPMSSA 180
|
|
|
QY 214  ROVSLL 219
|
|
|
Db 181  ROVSLL 186
|
|
|
RESULT 5
Q8RWPF6 PRELIMINARY; PRT; 186 AA.
AC Q8RWPF6;
DT 01-JUN-2002 (T-EMBLrel. 21, Created)
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
DE FRIGIDA protein (Fragment)
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Alguetrum;
RA Hagenblad J., Nordborg M.;
RT "Sequence Variation and Haplotype Structure Surrounding the Flowering
Time Locus FRI in Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY092684; AAM12137.1;
FT NON_TER 1
FT NON_TER 186
FT NON_TER 186
SQ SEQUENCE 186 AA; 20813 MW; 4643EF0A7ADDB7C0 CRC64;

Query Match 29.7%; Score 921; DB 10; Length 186;
Best Local Similarity 98.9%; Pred. No. 5.3e-57;
Matches 184; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 34  KIVETSTSDITIGSKQPFLLKSDLAARSVAVETFKRQFDLQKHIESIENAIISK 93
|
|
|
Db 1  KIVETSTSDITIGSKQPFLLKSDLAARSVAVETFKRQFDLQKHIESIENAIISK 60
|
|
|
QY 94  LEANGVVLARNNNFHQPMLSPPNNVSVETTTVVSQPSQEIIVPETSNNKPEGGRMCCLMC 153
|
|
|
Db 61  LEANGVVLARNNNFHQPMLSPPNNVSVETTTVVSQPSQEIIVPETSNNKPEGGRMCCLMC 120
|
|
|
QY 154  SKGLRKYYIANISDQAKLMEIIPSAKLAKPAKFVLDICIGKIFYLQGRRAFTKESPMSSA 213
|
|
|
Db 121  SKGLRKYYIANISDQAKLMEIIPSAKLAKPAKFVLDICIGKIFYLQGRRAFTKESPMSSA 180
|
|
|
QY 214  ROVSLL 219
|
|
|
Db 181  ROVSLL 186
|
|
|
RESULT 6
Q8RVCO PRELIMINARY; PRT; 138 AA.
AC Q8RVCO;
DT 01-JUN-2002 (T-EMBLrel. 21, Created)
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
DE FRIGIDA protein (Fragment)
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Various strains;
RA Hagenblad J., Nordborg M.;
RT "Sequence Variation and Haplotype Structure Surrounding the Flowering
Time Locus FRI in Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY092437; AAM12050.1;
DR EMBL; AY092438; AAM12051.1;
DR EMBL; AY092439; AAM12052.1;
DR EMBL; AY092440; AAM12053.1;
DR EMBL; AY092441; AAM12054.1;
DR EMBL; AY092442; AAM12055.1;
DR EMBL; AY092443; AAM12056.1;
DR EMBL; AY092444; AAM12057.1;
DR EMBL; AY092445; AAM12058.1;
DR EMBL; AY092446; AAM12059.1;
DR EMBL; AY092447; AAM12060.1;
DR EMBL; AY092448; AAM12061.1;
DR EMBL; AY092449; AAM12062.1;
DR EMBL; AY092450; AAM12063.1;
DR EMBL; AY092451; AAM12064.1;
DR EMBL; AY092452; AAM12065.1;
DR EMBL; AY092453; AAM12066.1;
DR EMBL; AY092454; AAM12067.1;
DR EMBL; AY092455; AAM12068.1;
DR EMBL; AY092456; AAM12069.1;
FT NON_TER 1
FT NON_TER 138
FT NON_TER 138
SQ SEQUENCE 138 AA; 15674 MW; 17910646CD19CEEB CRC64;

Query Match 22.5%; Score 697; DB 10; Length 138;
Best Local Similarity 100.0%; Pred. No. 1.8e-41;
Matches 138; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 370  AQSPLAFKEAATKQLAVLSSVMQCMETHKLPKELPGWIKQIVSLKDTLQDKEME 429
|
|
|
Db 1  AQSPLAFKEAATKQLAVLSSVMQCMETHKLPKELPGWIKQIVSLKDTLQDKEME 60
|
|
|
QY 430  EKARSLSLMEEAALAKRMYNQIKRPLSPMEMPVTSSSYSPYRDRSPFSQRDDQDE 489
|
|
|
Db 61  EKARSLSLMEEAALAKRMYNQIKRPLSPMEMPVTSSSYSPYRDRSPFSQRDDQDE 120
|
|
|
QY 490  ISALVSSYLGPTSTSPHR 507
|
|
|
Db 121  ISALVSSYLGPTSTSPHR 138
|
|
|
RESULT 7
Q8RV51 PRELIMINARY; PRT; 104 AA.
AC Q8RV51;
DT 01-JUN-2002 (T-EMBLrel. 21, Created)
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE FRIGIDA protein (Fragment)
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Various strains;
RA Hagenblad J., Nordborg M.;
RT "Sequence Variation and Haplotype Structure Surrounding the Flowering
Time Locus FRI in Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY092537; AAM12090.1;
DR EMBL; AY092538; AAM12091.1;
DR EMBL; AY092539; AAM12092.1;
DR EMBL; AY092540; AAM12093.1;
DR EMBL; AY092541; AAM12094.1;

```

DR EMBL; AY092542; AAM12095.1; -  
DR EMBL; AY092543; AAM12096.1; -  
DR EMBL; AY092544; AAM12097.1; -  
DR EMBL; AY092545; AAM12098.1; -  
DR EMBL; AY092546; AAM12099.1; -  
DR EMBL; AY092547; AAM12100.1; -  
DR EMBL; AY092548; AAM12101.1; -  
DR EMBL; AY092549; AAM12102.1; -  
DR EMBL; AY092550; AAM12103.1; -  
DR EMBL; AY092551; AAM12104.1; -  
DR EMBL; AY092552; AAM12105.1; -  
DR EMBL; AY092553; AAM12106.1; -  
DR EMBL; AY092554; AAM12107.1; -  
DR EMBL; AY092555; AAM12108.1; -  
DR EMBL; AY092556; AAM12109.1; -  
FT NON TER 1  
SQ SEQUENCE 104 AA; 35D92637P91C72F1 CRC64;  
  
Query Match 19.5%; Score 574; DB 10; Length 104;  
Best Local Similarity 100.0%; Pred. No. 5.3e-33;  
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 506 HRSRRSPYVPLPHGGGLGRSVYAYEHLAPNSYSPGHGHLRQYSPSLVHGQRHPLQYS 565  
Db 1 HRSRRSPYVPLPHGGGLGRSVYAYEHLAPNSYSPGHGHLRQYSPSLVHGQRHPLQYS 60  
  
QY 566 PPIHQOQLPYGIQVRYHSPSEERYGLSNQSPRNSLDPK 609  
Db 61 PPIHQOQLPYGIQVRYHSPSEERYGLSNQSPRNSLDPK 104  
  
RESULT 8  
Q8L7D6 PRELIMINARY; PRT; 558 AA.  
AC Q8L7D6; ID Q8L7D6; PRELIMINARY; PRT; 532 AA.  
DT 01-OCT-2002 (TRENBLrel. 22, Created)  
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)  
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
DE Hypothetical protein.  
GN AT5G48390.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eucosids II; Brassicales; Brassicaceae; Arabidopsi.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Southwick A., Nguyen M., Tripp M., Palm C.J., Jones T., Wu T.,  
RA Carninci P., Chen H., Cheuk R., Chan M.M., Chang C.H., Dale J.M.,  
RA Deng J.M., Hayashizaki Y., Hsuan V.W., Lee J.M., Ishida J., Kamiya A.,  
RA Kawai J., Kim C.J., Narusaka M., Quach H.L., Sakurai T., Satou M.,  
RA Seki M., Shinn P., Tang C.C., Toroumi M., Wallender E.K., Wong C.,  
RA Wu H.C., Yamada K., Yu G., Yuan S., Shinozaki K., Ecker J.,  
RA Theologis A., Davis R.W.;  
RA Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Nguyen M., Karlin-Neumann G., Southwick A., Tripp M., Miranda M.,  
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,  
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
RA Ecker J., Theologis A., Davis R.W.;  
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY136326; AAM36992.1; -  
DR EMBL; BT000104; AAN15423.1; -  
SQ Hypothetical protein.  
KW SEQUENCE 558 AA; 62616 MW; 418ASBAC1CF942DC CRC64;  
  
Query Match 11.0%; Score 340; DB 10; Length 558;  
Best Local Similarity 24.9%; Pred. No. 1.6e-15;  
Matches 139; Conservative 94; Mismatches 223; Indels 102; Gaps 22;

QY 23 HQSPRRRLPKIVETESTSMDITIGSQKQPOFLKSIDELAAFSVAVET-FKRQDDDLQK 81  
Db HELEDQKE-----YEIKTR-----XAELEKKGAVEAKAALERLQK 99  
  
QY 82 HIESIENAIKSLGNGVLAARNNHOPMLSP-----PRN-----NVSVEITVTVS 129  
Db KRDAAMFTINSALD-----KYNN--APVSKPSYGERWPNQNAVEDSSNVFAADSIIDD 149  
  
QY 130 QP-----SQBIVP-----ETSNKPEGRMCELMCSKGLRKYIVANISDAKLMETPSAL 179  
Db NPDGIVQDVQISPMVNGYEVKAYPQLLKCGDMDSITGLHKFVSDNRKNLASKEIPIPAF 209  
  
QY 180 KLAKEPAKFVLDICIGFY-----LQGRRAFTKESPMSSARQVSLILIE--SFLMP-DR 230  
Db RAAANPASLVDSLEGFYPMFAPTADGK---KDALLGMRRTCTIMLECLSIILSGLDR 265  
  
QY 231 GKGVKTESWIKDEAETAAYAVAKRLMT-EGGLAAAEKMDARGLLILLVACFGVPNSRST 289  
Db NCLAVLSQNVKRAKTIAGWNPILLESMDACNGNSLENAFQLQLATPAIVADFKED 325  
  
QY 290 DLLDLIMSGNSIAGALKRQSLFVPMVSGIVESSIKRGMHIEALEMYVTFGMEKFSAA 349  
Db ELLKLIIPWRRRRQAALCTSLGLAEKMPGVIEVLVNSGKQIDAVNLAFAPFELTEQFSVP 385  
  
QY 350 LVLTSLFKMSKESFERAKRKAQSLAPFAKAAATKQALVLSVMQCMETHKLDPAKELPGWQ 409  
Db SLKSYLIEARRSPQGRPNASPAVQDEFNRELIGLTKVINCIEHSL-EEQP--- 440  
  
QY 410 IKEQIVSLEKDTLQDKEMBEKARSLMEEAALAKRMYNQIKRPLSPMEMPVPTSS 469  
Db VEPHKLRIQLBEAKADKKRATPKM-----PQKPRGA---QPRVTDNN 483  
  
QY 470 -----YSPIYDRSPSPSDDDDQDEISALVSSYLGPSTSFHRSRSPSEYVWPLPH 520  
Db NNINNTGYGRVPER-YPQYVYDNRFLSGPIMAAQPPPPPPPTTYTFNP---AP-AH 538  
  
QY 521 GGLGRSVYAYEHLAPNSY 538  
Db GNFYANCYQYQAPPPPPY 556  
  
RESULT 9  
Q9LUV4  
ID Q9LUV4; PRELIMINARY; PRT; 532 AA.  
AC Q9LUV4;  
DT 01-OCT-2000 (TRENBLrel. 15, Created)  
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)  
DE Putative hydroxyproline-rich glycoprotein.  
GN MCB17.20 OR AT3G22440.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eucosids II; Brassicales; Brassicaceae; Arabidopsi.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Columbia;  
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;  
RA Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Columbia;  
RX MEDLINE=20277480; PubMed=10819329;  
RA Nakamura Y.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence  
RT features of the regions of 4,504,864 bp covered by sixty pl and TAC  
RT clones.";  
RL DNA Res. 7:131-135(2000).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Liu S.X., Pham P.K., Banh J., Dale J.M., Goldsmith A.D.,  
RA Jiang P.X., Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M.,

[illegible]

DB 205 FLHLVAFAELGSLFTEBICDVFILSKYKQATTCICKIGLDRNRIGVLVQKFLDTGRLL 264  
QY 332 EALEMVTTFMEDKFSALVLTSLFKWSKESFERAKKAQSPAFK-EAATKOLAVLSSV 390  
DB 265 VAIRIYENWGEFEPVSIKLSKNSREAKKVCAGNYSLVKQNEATDKLSALRAV 324  
QY 391 MCMETHKLDPAKELPGQIKQIIVLSLEKTLQDQKEMEBEAKSLSLMEBAALAKMNYQ 450  
DB 325 IKVVEKNIES-----EFMEE-KLEECVKELEDQAKQKKA-TKFNSPANFPQPOEQ 374  
QY 451 QI--KRPRL---SPME-----MPEVTSSSYSPYDRSPSPQDDQDEISALVSSYLGPS 501  
DB 375 KYDNKRPVANGSMYENLTTPPLRPOQPPU-----LTPSQILQNPVGLSSIL-PG 428  
QY 502 TSFPHRSRSPYVPLPHGGLGSRVAYAEHLAPNSYSPGHGHLRQYSPSLVHQRP 561  
DB 429 VAVPYGNFRALPGSVPAF--ASRPVFVQ-----QTGVG-----MPP 463  
QY 562 LQYSPPIHQ 571  
DB 464 PQYRPPYPQ 473  
  
RESULT 11  
ID O04649 PRELIMINARY; PRT; 862 AA.  
AC O04649;  
DT 01-JUL-1997 (TremBLrel. 04, Created)  
DT 01-JUL-1997 (TremBLrel. 04, Last sequence update)  
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)  
DE A.TW021B04.8 protein.  
GN A.TW021B04.8  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
OX NCBI\_TaxID=3702;  
[1]  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RA Dante M., Wamsley P., Gibson A.;  
RT "The sequence of A. thaliana TW021B04.";  
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
[2]  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RA Wash-U;  
RT "The A. thaliana Genome Sequencing Project.";  
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
[3]  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RA Waterston R.;  
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF007271; AB561078.1; -- A83368FP0C6D4F31 CRC64;  
SQ SEQUENCE 862 AA; 96960 MW; A83368FP0C6D4F31 CRC64;  
  
Query Match 9.2%; Score 286.5; DB 10; Length 862;  
Best Local Similarity 22.1%; Pred. No. 1.7e-11;  
Matches 132; Conservative 119; Mismatches 231; Indels 115; Gaps 21;  
  
QY 18 PLLRQSQRRRLPKIVETESTSDITIGSKQPOFLKSIDELAASVAVETFKQFD 77  
DB 355 FYEQKHVIR-----PRLTEPTAPSONLTVCFNMKHELTSMNLSIF-VFGETLRHKF- 407  
QY 78 DLQKHIESIENADSKLEBNGVVLARNNN-----PQPMSPPRNNVSVETTVTSQ 130  
DB 408 -----NQNIFFLLFKLKQPEV-----VCVPTGKQVKE 434  
QY 131 PSQEIPTSNKEGGRMELMCSKG-----LRKIIVANISDQAKMEEFSAKLKAKE 184  
DB 435 SGADHQPDTHATPSGTETKILNLSGSIKADMLRELVE---KQPLKESEDLSNALKCTPD 491

QY 185 PAKFVLDCIKFYLOQRRAFTKESPMSSARQVSLILLESFILMPDRGKGVKIESWIKDE 244  
DB 492 PAKFLDTSMALCPTTEGGYEFKMLITSASCSLLNLQKLLP-----KIGHPVKGD 544  
QY 245 AETAAVARERLMTTEGLAAAEKMDARGLLLVACFGVPSNFRSTDLDDLIRMSGSEIA 304  
DB 545 AKKLVATWKDKI-----AKSKDQLEVICFLOFGIFGVSEFKADLLGLLONSQWQTVS 600  
QY 305 GALKRSOFLVPMVSGIVSESIKRGMIHIALEVMYTFGMEDKTS-AALVLTSLFKWSKESF 363  
DB 601 PDLCOFLGLDPAIPGFTQNLKTHRIKAIKDIYISFGMVHRFPQVSAIINDSLR-ITKESA 660  
QY 364 ERACKAQSPLAFKEAA-TKQAVLSSVMQCMETHKLDPAKELPGWQIKQIIVLSLEKDTL 422  
DB 661 EKSYREAKNSETTQVAIDQVRAALRAAKICSCHESEFQJ--GDLEQIKSLK--L 716  
QY 423 QLDKEMEBEAKSLSLMEBAALAKMNYNQIKRPLSPMEMPPVTSSSYSPYDRSPSPQ 482  
DB 717 RRTNSNGSGSGSASSKPDSTIKQ---SQTAKPPTVA-EVAPVTS-----NIPLE 761  
QY 483 RDDQDEISALVSSYLGPSSTSPHRSRSPYVPLPHGGLGRSV---YAYEHLA--PN 536  
DB 762 PSTEAASSA-----SKPFSKONR-----GKRSMSGNQQSGHLSHTS 802  
QY 537 SYSPGHGHLRHOYSPSLVHGQRHPLQYSPPIHQQQ--LPYGIQVRVYRHSPEERY 591  
DB 803 NHYPSHDYSLNQRLTWPVDNYDRGFTGFPNPDYNNQWQGPQGFYHLYQPLDPY 859  
  
RESULT 12  
ID Q940H8 PRELIMINARY; PRT; 532 AA.  
AC Q940H8;  
DT 01-DEC-2001 (TremBLrel. 19, Created)  
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)  
DT 01-OCT-2002 (TremBLrel. 22, Last annotation update)  
DE Hydroxyproline-rich glycoprotein-like protein.  
GN Z97337.18.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
OX NCBI\_TaxID=3702;  
[1]  
RN  
RP SEQUENCE FROM N.A.  
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,  
Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,  
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
Ecker J., Theologis A., Davis R.W.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
[2]  
RN  
RP SEQUENCE FROM N.A.  
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,  
Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,  
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
Ecker J., Theologis A., Davis R.W.;  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
[3]  
RN  
RP SEQUENCE FROM N.A.  
RA Haas B.U., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,  
Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;  
RT "Full-length messenger RNA sequences greatly improve genome  
annotation.";  
RN Genome Biol. 0:0-0(2002).  
[4]  
RN  
RP SEQUENCE FROM N.A.  
RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,  
Feldmann K.



RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Kotani H.,  
RA Tabata S.,  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence  
RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC  
RT clones";  
RL DNA Res. 7:31-63(2000).  
DR EMBL: AB020745; BAA96959.1; -;  
DR InterPro: IPR001440; TPR.  
DR Pfam: PF00515; TPR; 2.  
SQ SEQUENCE 1337 AA; 148858 MW; 97A476B4CB3EC6B3 CRC64;  
  
Query Match 8.5%; Score 262.5; DB 10; Length 1337;  
Best Local Similarity 23.7%; Pred. No. 1.5e-09;  
Matches 121; Conservative 70; Mismatches 169; Indels 151; Gaps 21;  
  
QY 37 ETESTS--WDITIGSKQPFKLSIDELAA-----ESVAVETFKQPF-- 76  
Db 3 DTRSVASLMDST--SSKIQQLQKAPAELESQRAVTLNLKWLKEEFHGHGSLKRRFHE 60  
QY 77 -DDLQKHIESIENADSKLENGVVLARN-----NNFHQ 110  
Db 61 LEDQKEVETKRAQELLEKKAAVEAKEAALERLOKXEDAMFTINSALDKYNN--A 118  
QY 111 PMLSP-----PRN-----NVSVETTVTSQP-----SSEIVP-----ETSNNKPGGRM 148  
Db 119 PYSKPSVGERWPQNAVEDSSNVFAADSIITDDNPDGIQVQISPMWNGYEVKAYPOLKLL 178  
QY 149 CELMCSKGLKRYIYANISDQAKLMEIIPSAKLAKPAKFLVDCIGKPY-----LQGR 202  
Db 179 CDMOSTGHLKRVSDNRKNVLSLKEEIPMAFRAANPASVLDSLEGFYFWAPTADGK- 237  
QY 203 APTKESPSSARQVSLILE--SFLMP--DRGKGVKIESWIKDEAETAATAVAVRKLMT- 258  
Db 238 ---KDALNLLGMRRTICMLMECLISLLGLDNCIAVLVLSQVNHAKRTIAEGWNPILLES 294  
QY 259 EGGLEAAEKMDARGILLVACFGVPSNFRSTDLIDLIMSGNIEIAGALKRSQFLPMVS 318  
Db 295 DMDACNGSLEAHAFLOLATAIVADVDFEDELKLIIPMVSRRRQAAELCRSLGLAEKMP 354  
QY 319 GIVESISIRKRGHIEALEMVTYFGMEDKFSAAALVLTSLFKMSKESFERAKRAQSPLAFKE 378  
Db 355 GVTEVLVNSGRQIDA-----DEFN----- 373  
QY 379 AATKQALVSSVMQCMETHKLDPAKELPGWIKQIIVSLEKOTILDKEMEEKARSLSLM 438  
Db 374 --ERELIGLTKVTKIEHSLE--EQVP-----VEPLHKLQLEKAKADKKRATEPM 422  
QY 439 EBAALAKRWYNOQIKRPLSPWMPPTVSSS 469  
Db 423 K-----PQKPRGA---QPRVTDNN 440  
  
RESULT 15  
Q9FFP1 PRELIMINARY; PRT; 470 AA.  
AC Q9FFP1;  
DT 01-WAR-2001 (TREMELrel. 16, Created)  
DT 01-WAR-2001 (TREMELrel. 16, Last sequence update)  
DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)  
DE Genomic DNA, chromosome 5, P1 clone:NQK4 (AR5G16320/NQK4\_4).  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC euroside 11; Brassicales; Brassicaceae; Arabidopsi.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Columbia;  
RX MEDLINE=97471969; PubMed=9330910;  
RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,  
RA Miyajima N., Tabata S.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence  
RT features of the 1.6 Mb regions covered by twenty physically assigned

RT P1 clones";  
RL DNA Res. 4:215-230(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Shinn P., Chen H., Cheuk R., Kim C.J., Meyers M.C., Banh J.,  
RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,  
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,  
RA Kawai J., Lam B., Lee J.M., Lin J., Mirand M., Narusaka M., Onodera C.S.,  
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,  
RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,  
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,  
RA Ecker J.R.;  
RT "Arabidopsis cdna clones";  
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Cheuk R., Chen H., Kim C.J., Shinn P., Banh J., Bowser L.,  
RA Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,  
RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,  
RA Lee J.M., Lin J., Mirand M., Narusaka M., Nguyen M., Onodera C.S.,  
RA Palm C.J., Quach H.L., Sakurai T., Satou M., Southwick A.,  
RA Tang C.C., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yu S.,  
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;  
RT "Arabidopsis ORF clones";  
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB005242; BAB09599.1; -;  
DR EMBL: AY075663; AAL77670.1; -;  
DR EMBL: AY101525; AAM26646.1; -;  
SQ SEQUENCE 470 AA; 52831 MW; 2924997BB01366D6 CRC64;  
  
Query Match 7.9%; Score 246; DB 10; Length 470;  
Best Local Similarity 22.5%; Pred. No. 4.9e-09;  
Matches 125; Conservative 92; Mismatches 204; Indels 134; Gaps 22;  
  
QY 56 LKSDLEAASVAVETPKROFDLQKHIESIE-----NAIDSKLENGVVLAAARNNF 108  
Db 11 INQIDE-----KKEUKKAFDLDQHRSLSPSLSWSEIDSHFSSLOSLAGRFL 64  
QY 109 HQPMLSPRRNVSVETTVT---VSQPSQIVPETSINKPEGRCMELCMCSGLKRYIYANI 165  
Db 65 HS---TSFLEHDSYRIDASDAGKSSSESEVSEQVPELRLALCEKIDIGLIKYLIRIW 121  
QY 166 SDQAKLMEIIPSAKLAKPAKFLVDCI--GKFLQGRRAFTKESPMS-----ARQVSL 219  
Db 122 DDETPLNQVESAIRVSPDTASWLVDAIEGSNV-----TPSSGRSPDVRVFL 171  
QY 220 LLESFLMPDRGKGVKIESWIKDEAETAATAVAVRKLMTTEGGLAAAEKMDARGILLVAC 279  
Db 172 LMEVLIEI---NANITVDT--RNRKALAYHWKSV-----GVKPEALVFLHLVAA 218  
QY 280 FGVPNSFRSTDLIDLIMSGS-----NETAGALKRSQFLVPMVSGIVESISIRGMHI 331  
Db 219 FELGSEFDELSGDYVFMIAKYQATLVNCNKIGVDRKR-----VGLIKTLDSGKPI 271  
QY 332 EALEMVTYFGMEDKFSAAALVLTSLFKMSKESPERA-----KRKAQSPFLAFKEAATKQLA 385  
Db 272 LAVKFMVECGMTDEFEPFIPVKSVIKDCREAAALRVCEVDNYSLSQN-----EASDKEVS 326  
QY 386 VLSSVMQCMETHKLDPAKELPGWIKQIIVSLEKOTILDKEMEEKARSLSLMEEAALAK 445  
Db 327 ALKPLIKIIXDQNLB--SEPTQEKVEERVEELEKN-----KALKK 364  
QY 446 RMVNOQIKRPLSPWMPPTVSSSYSPYVDRSFRSQRDDQDEISALVSSYLGSTSPFP 505  
Db 365 R---NTTNPKQFPQKQKRT-----RDCNKGSOVPVPVPSQQLLSRPEALLMEHS-- 412  
QY 506 HRSRSPSYWVPLPHGGLGRSVY-----AYEHLAPNSYSPGHGHLRHRYQSYSLVHGQRH 560  
Db 413 -----HRGLQNLNPLYGLMTSAFSGVYVNVNPLTGLFGSGATPQ---SLYAAQQT 455  
QY 561 ---PLOSYPPIHQ 571  
Db 456 GYVLPPQYHPPIYSQ 470

Search completed: August 14, 2003, 10:40:00  
Job time : 110 secs

PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160989.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161359.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0145386.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146388.	PR	28-OCT-1999;	99US-0161920.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161992.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147204.	PR	29-OCT-1999;	99US-0162142.
PR	04-AUG-1999;	99US-0147302.			
PR	05-AUG-1999;	99US-0147192.			
PR	05-AUG-1999;	99US-0147260.			
PR	06-AUG-1999;	99US-0147303.			
PR	06-AUG-1999;	99US-0147416.			
PR	09-AUG-1999;	99US-0147493.			
PR	09-AUG-1999;	99US-0147935.			
PR	10-AUG-1999;	99US-0148171.			
PR	11-AUG-1999;	99US-0148319.			
PR	12-AUG-1999;	99US-0148341.			
PR	13-AUG-1999;	99US-0148565.			
PR	13-AUG-1999;	99US-0148684.			
PR	16-AUG-1999;	99US-0149368.			
PR	17-AUG-1999;	99US-0149175.			
PR	18-AUG-1999;	99US-0149426.			
PR	20-AUG-1999;	99US-0149722.			
PR	20-AUG-1999;	99US-0149723.			
PR	20-AUG-1999;	99US-0149929.			
PR	23-AUG-1999;	99US-0149902.			
PR	23-AUG-1999;	99US-0149930.			
PR	25-AUG-1999;	99US-0150566.			
PR	26-AUG-1999;	99US-0150884.			
PR	27-AUG-1999;	99US-0151065.			
PR	27-AUG-1999;	99US-0151066.			
PR	27-AUG-1999;	99US-0151080.			
PR	30-AUG-1999;	99US-0151303.			
PR	31-AUG-1999;	99US-0151438.			
PR	01-SEP-1999;	99US-0151930.			
PR	07-SEP-1999;	99US-0152363.			
PR	10-SEP-1999;	99US-0153070.			
PR	13-SEP-1999;	99US-0153758.			
PR	15-SEP-1999;	99US-0154018.			
PR	16-SEP-1999;	99US-0154039.			
PR	20-SEP-1999;	99US-0154779.			
PR	22-SEP-1999;	99US-0155139.			
PR	23-SEP-1999;	99US-0155486.			
PR	24-SEP-1999;	99US-0155659.			
PR	28-SEP-1999;	99US-0156458.			
PR	29-SEP-1999;	99US-0156596.			
PR	04-OCT-1999;	99US-0157117.			
PR	05-OCT-1999;	99US-0157753.			
PR	06-OCT-1999;	99US-0157865.			
PR	07-OCT-1999;	99US-0158029.			
PR	12-OCT-1999;	99US-0158232.			
PR	13-OCT-1999;	99US-0158369.			
PR	13-OCT-1999;	99US-0159293.			
PR	13-OCT-1999;	99US-0159294.			
PR	13-OCT-1999;	99US-0159295.			
PR	14-OCT-1999;	99US-0159329.			
PR	14-OCT-1999;	99US-0159330.			
PR	14-OCT-1999;	99US-0159331.			
PR	14-OCT-1999;	99US-0159637.			
PR	14-OCT-1999;	99US-0159638.			
PR	18-OCT-1999;	99US-0159584.			
Query Match 10.8%; Score 336; DB 21; Length 454;			Query Match 10.8%; Score 336; DB 21; Length 454;		
Best Local Similarity 25.5%; Pred. No. 2.4e-21;			Best Local Similarity 25.5%; Pred. No. 2.4e-21;		
Matches 126; Conservative 84; Mismatches 199; Indels 86; Gaps 19;			Matches 126; Conservative 84; Mismatches 199; Indels 86; Gaps 19;		
QY	85	SIENAIKSKLESXGVLAARNNNFHQFMILSP-----PRN-----NVSVETTVTSOP- 131	QY	85	SIENAIKSKLESXGVLAARNNNFHQFMILSP-----PRN-----NVSVETTVTSOP- 131
Db	3	TINSALD-----KYNN--APVSKPSYGERWPQNAVEDSSNVFAADSIIDNDPD 48	Db	3	TINSALD-----KYNN--APVSKPSYGERWPQNAVEDSSNVFAADSIIDNDPD 48
QY	132	-----SQEIVP-----ETSNKEPGRMCELMCKGLKRYIYANISDQAKLMEIIPGALKLA 182	QY	132	-----SQEIVP-----ETSNKEPGRMCELMCKGLKRYIYANISDQAKLMEIIPGALKLA 182
Db	49	GIVODVQLSPVMGNYEVKAYPQLLKCGDMDSTGLHKFVSDNRKNLASKEEIPAFRAA 108	Db	49	GIVODVQLSPVMGNYEVKAYPQLLKCGDMDSTGLHKFVSDNRKNLASKEEIPAFRAA 108
QY	183	KEPAKFVLDICIGFY-----LQGRAFTKESPMSSARQVSLILE--SFLMP-DRGKG 233	QY	183	KEPAKFVLDICIGFY-----LQGRAFTKESPMSSARQVSLILE--SFLMP-DRGKG 233
Db	109	ANPASLVLDLSLEGFYPMPEATDGK---KDNLLGMRTICIMLMECLISLLSGLDRNSL 164	Db	109	ANPASLVLDLSLEGFYPMPEATDGK---KDNLLGMRTICIMLMECLISLLSGLDRNSL 164
QY	234	KVIESWIKDEAETRAVAVWKRMT-EGGLAAAKMDARGLLLVACFGVPSNFRSTDLL 292	QY	234	KVIESWIKDEAETRAVAVWKRMT-EGGLAAAKMDARGLLLVACFGVPSNFRSTDLL 292
Db	165	AVLSQNVKRAKTIAGWNPLLESMDACNGNSLEAHAFLOLLATFAIVADFKEDELL 224	Db	165	AVLSQNVKRAKTIAGWNPLLESMDACNGNSLEAHAFLOLLATFAIVADFKEDELL 224
QY	293	DLIRMSGNSIAGALKRSQFLVPMVSGIVESSIKRGMHTEALEMVYTFGMEDKFSAAVL 352	QY	293	DLIRMSGNSIAGALKRSQFLVPMVSGIVESSIKRGMHTEALEMVYTFGMEDKFSAAVL 352
Db	225	KLIPWVSRRRQAEELCSLGLAEKMPGVIEVLVNSQKIDAVNAFAFELTQSPVSILL 284	Db	225	KLIPWVSRRRQAEELCSLGLAEKMPGVIEVLVNSQKIDAVNAFAFELTQSPVSILL 284
QY	353	TSFLMSKESFERAKKAQSPFAKKAATQKQAVLSVNCQCMETHKLDPAKELPWQIKE 412	QY	353	TSFLMSKESFERAKKAQSPFAKKAATQKQAVLSVNCQCMETHKLDPAKELPWQIKE 412
Db	285	KSYLIEARRSSPOGRPGNASPAVDSEFNERELIGLTKVICKIEEHSLE--EQYP----- 336	Db	285	KSYLIEARRSSPOGRPGNASPAVDSEFNERELIGLTKVICKIEEHSLE--EQYP----- 336
QY	413	QIVSLEKDTLQDKEMEEKARSLSLMEEAALAKRMYNQOIKRPLSPMBMPVTSSS--- 469	QY	413	QIVSLEKDTLQDKEMEEKARSLSLMEEAALAKRMYNQOIKRPLSPMBMPVTSSS--- 469
Db	337	-VEPLHRIIQLQEKAKADKKRATPEMK-----POKPRGA---QPRVTNNNNI 382	Db	337	-VEPLHRIIQLQEKAKADKKRATPEMK-----POKPRGA---QPRVTNNNNI 382
QY	470	-----YSPYRDRSPSQDDQDEISALVSSYLGPSSTSPHRRRRSPYVPLPHGL 523	QY	470	-----YSPYRDRSPSQDDQDEISALVSSYLGPSSTSPHRRRRSPYVPLPHGL 523
Db	383	NNKTYGVRVIPER-YPOYYVDRNRPFLSGPIMAAQPPPPPPPTVTENP---AP-AHGNF 437	Db	383	NNKTYGVRVIPER-YPOYYVDRNRPFLSGPIMAAQPPPPPPPTVTENP---AP-AHGNF 437
QY	524	GRSVYAYEHLAPNSY 538	QY	524	GRSVYAYEHLAPNSY 538
Db	438	YANCYQCAPPPPPY 452	Db	438	YANCYQCAPPPPPY 452
RESULT 7			RESULT 7		
AAG39609			AAG39609		
ID	AAG39609 standard; Protein; 456 AA.		ID	AAG39609 standard; Protein; 456 AA.	
XX	AAG39609;		XX	AAG39609;	
DT	18-OCT-2000 (first entry)		DT	18-OCT-2000 (first entry)	
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 49034.		XX	Arabidopsis thaliana protein fragment SEQ ID NO: 49034.	
DE	Protein identification; signal transduction pathway; metabolic pathway;		DE	Protein identification; signal transduction pathway; metabolic pathway;	
XX	hybridisation assay; genetic mapping; gene expression control; promoter;		XX	hybridisation assay; genetic mapping; gene expression control; promoter;	
KW			KW		

Db 90 KDAAMFTINSALD-----KYN--APVSKSVGERWPCNAVESNNVFAADSTDD 139  
QY 130 QP-----SDEIVP-----ETSKPGRCMELMCKGLRKVIYANI SDOAKLMEEIPESAL 179  
Db 140 NPDGIVQDVQISPVNVCNVEYKAYPQLLXLCGDMDSGLHFKFVSDNRKNLASLKEEIPVAF 199  
QY 180 KLAKEPAKVLDCIGKFY-----LQRRATKESPMGSAQVSLILLE--SFLIMP-DR 230  
Db 200 RAAANPASLVLSLGEFYPMAPTADGK-----KQANLGMERTCIMLMECLISILLGLDR 255  
QY 231 GKGKVKIESWIDEAETAAVAKELMT-EGGLAAAEKMDARGILLIVACFGVPSNFRST 289  
Db 256 NCLAVVLSQNVKRAKTIAGWNPLLESILDMDACNGNSLEAHAFIQLLATPAIIVADFKED 315  
QY 290 DILLDIRMGSSNEIACALRQSFLPMVSGIVESSIKRGMHIEALEMVIYFGMEDKFSAA 349  
Db 316 ELIKLIPMYRRROAELCRSLGAEKMPGVIEVUNSGKIDAVNALAFAPFELTEQSPV 375  
QY 350 LVLTSPFKMSKESFERAKRKAQSLP--AFKEAATKOLAVLSVMQCMETHKLDPAKPLG 407  
Db 376 SLLKSVLIARSSPQGRPNASPAVQASDEFNERELGLKTVIKCIEHSLE--EQVP- 432  
QY 408 WQKEQIVLEKDTLQDKEMEBEKARSLSLMEEAALAKRMYNQOKRPRISPMEMPPVTS 467  
Db 433 -----VEPLHKRIILQLEKAKADKRATEPMK-----POPKRPGA--QPRVTD 473  
QY 468 SS-----YPIYDRSFSQDDDDDETSAVSSVLGPGSTSFPHRSRRSPFYWVPL 518  
Db 474 NNNNNNTGTVGRVIPER-IPQYVYDNRPFUSGPFMAAQPPPPPTOTYTFNP---AP- 528  
QY 519 PHGGLGRSVYAYEHLAPNSY 538  
Db 529 AHGNFYANCYQYQAPPPPY 548  
RESULT 6  
AAG06059  
ID AAG06059 standard; Protein; 454 AA.  
XX  
AC AAG06059;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 2693.  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EF1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 16-APR-1999; 99US-0128714.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134258.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137282.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.

PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 24-JUN-1999; 99US-0140823.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 08-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 10.9%; Score 337; DB 21; Length 550;

Best Local Similarity 25.2%; Pred.No. 2.7e-21;

Matches 141; Conservative 93; Mismatches 222; Indels 104; Gaps 23;

QY 23 HQSEORRELPKIVETESTSMDITIGSQOPFLKSIDELAAFSAVET-FKQFDDLOK 81

Db 49 HELEDQKE---YETKTR-----KQELLEKKKAVERAKEKALRELQK 89

QY 82 HIESIENAIKSLKESNGVLAARNNNHQMPLSP-----PRN-----NVSVETTUTVS 129



PR 25-FEB-1999; 99US-011825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145951.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148584.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.

PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Watch	11.1%;	Score 346; DB 21; Length 548;
Best Local Similarity	25.1%;	Pred. No. 4e-22;
Matches 140;	Conservative	93; Mismatches 223; Indels 102; Gaps 22;
QY	23 HQSEQRRELEPKIVETESTSMDITIQQSKQPOPLKSIDELAASFVAVET-PKRQFDDLQK 81	
DB	49 HELEDQEKE---YETNTR-----KAQELLEKKAAVEAKEKAALERLQK 89	
QY	82 HIESTENALDSKLESNGVLAAARNNHQHMLSP-----PEN-----NVSVETTTVTS 129	
DB	90 KDAAMFTINSALD-----KYNN--APVKSPSGVERWQPONAVEDSNVFAADSIIDD 139	
QY	130 QP-----SQEITVP-----ETSNKPEGRMCMLCKSLGRLKYIYANISQAKLMEIPSA 179	
DB	140 NPDGIVQDVQIISPVMGNVEYKAYPQLLKCGDMDSITGLHKFYSDNRKNLASLKEEIPMAF 199	
QY	180 KLAKEPKXFLDCTGKY-----LQRRAFTHKESPMSSARQVSLLILE--SFLIMP-DR 230	
DB	200 RAAANPASLVLDSEGFYPMEAPTADGK----KDANLGMRETCTIMLECLISLLGLDR 255	
QY	231 GKGYKIBSWTKDAETAATAVAWRKLMWT-EGGLAAAEKMDARGLLLVACFGVPSNFRST 289	
DB	256 NSLAVLVQNKHRAKIAGWNPLIESLMDACNGNSLEAHAFLOLATPAIVADFKE 315	
QY	290 DLLDLIRMSGNEIAGALKRSQFLPVMVGIVESSIKRGMHIEALEMYYTTGEMEDKFSAA 349	
DB	316 ELLKLIPIMVSRROAELCSRLGLAEKMPGVIEVLVNSGKQIDAVNLAPAFELTEQFSVP 375	
QY	350 LVLTSLFKWSKESFERAKRKAQSPLAFKZAATKQLAVLSSVMQCWETHKLPAPAKELPGW 409	
DB	376 SLTSLYLTEARRSPQGRPNASPAVDNFNERELIGLKTIVIKCIEHSLE--EQYP--- 430	
QY	410 IKEQIVLSLEKDTLOLDKEMEBEKARSLSLMEERAAKRMYNQOIKRPILSPMEMPPVTSS 469	
DB	431 ----VEPLHKRIILQLEKAKADKKRATEPMK-----PQKPGRGA---QPRVTDNN 473	
QY	470 -----YSPIYDRSFSPDDDCDEISALVSSYLGPFSTSPHRSRSPRYMVPLPH 520	
DB	474 NNINNNKTGYGRVTPER-YFYQYVDNRNFLSGPIMAAQPPPPPQTTFNP---AP-AH 528	
QY	521 GGLGRSVYAVEHLAPNGV 538	
DB	529 GNFIANCYQYQAPPY 546	
RESULT 4		
AAG39607		
ID	AAG39607 standard; Protein; 560 AA.	
XX		
AC	AAG39607;	
XX		
DT	18-OCT-2000 (first entry)	
XX		
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 49032.	
XX		
KW	Protein identification; signal transduction pathway; metabolic pathway;	
XW	hybridisation assay; genetic mapping; gene expression control; promoter;	
XW	termination sequence.	
OS	Arabidopsis thaliana.	
XX		
PN	EP1033405-A2.	
XX		
PD	06-SEP-2000.	
XX		
XX	25-FEB-2000; 2000EP-0301439.	
XX		



PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	99US-0138899.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	99US-0140354.	PR	01-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	99US-0140695.	PR	07-SEP-1999;	99US-0152363.
PR	28-JUN-1999;	99US-0140823.	PR	10-SEP-1999;	99US-0153070.
PR	29-JUN-1999;	99US-0140991.	PR	13-SEP-1999;	99US-0153758.
PR	30-JUN-1999;	99US-0141287.	PR	15-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	99US-0141842.	PR	16-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	99US-0142154.	PR	20-SEP-1999;	99US-0154779.
PR	02-JUL-1999;	99US-0142055.	PR	22-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142390.	PR	23-SEP-1999;	99US-0155486.
PR	08-JUL-1999;	99US-0142803.	PR	24-SEP-1999;	99US-0155659.
PR	09-JUL-1999;	99US-0142920.	PR	28-SEP-1999;	99US-0156458.
PR	12-JUL-1999;	99US-0143977.	PR	29-SEP-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143542.	PR	04-OCT-1999;	99US-0157117.
PR	14-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	12-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144325.	PR	18-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144884.	PR	14-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144814.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160989.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161359.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146388.	PR	28-OCT-1999;	99US-0161920.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161992.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147204.	PR	29-OCT-1999;	99US-0162142.
PR	04-AUG-1999;	99US-0147302.			
PR	05-AUG-1999;	99US-0147192.			
PR	05-AUG-1999;	99US-0147260.			
PR	06-AUG-1999;	99US-0147303.			
PR	06-AUG-1999;	99US-0147416.			
PR	09-AUG-1999;	99US-0147493.			
PR	09-AUG-1999;	99US-0147935.			
PR</					

the flowering time of a plant -  
 Claim 30; Fig 6; 73pp; English.  
 The present sequence represents a polypeptide capable of specifically altering the flowering time of a plant. The polypeptide is encoded by the FRI (one locus-FRIGIDA) locus of Arabidopsis. The FRI polynucleotide is used to transform plants, so that the flowering time of a plant is altered. This is used, for example, for plants in which the leaves or tubers are a commercial product, where it is desirable to avoid 'blotting' (initiation of flowers and stem elongation) at too early a stage. Conversely, it may be desirable to alter flowering under certain circumstances e.g. to vary flower production across the seasons.

XX Sequence 609 AA;  
 Query Match 100.0%; Score 3104; DB 21; Length 609;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-272;  
 Matches 609; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNYPPTVAAPTATTANPLQRHSEORRELPKIVETESTSMDITIGOSKQFQFKSID 60  
 DB 1 MSNYPPTVAAPTATTANPLQRHSEORRELPKIVETESTSMDITIGOSKQFQFKSID 60

QY 61 ELAASFVAVFETFKQFDLOKHIESIENADSKLENGVLAARNNHQPMLSPPRNIV 120  
 DB 61 ELAASFVAVFETFKQFDLOKHIESIENADSKLENGVLAARNNHQPMLSPPRNIV 120

QY 121 SVETTVTVSQSQIIVPETSNKPEGGRMELMCKGLRKVIYANISDQAKLMEETPSALK 180  
 DB 121 SVETTVTVSQSQIIVPETSNKPEGGRMELMCKGLRKVIYANISDQAKLMEETPSALK 180

QY 181 LAKSPAKFVLDICIKFYLOQRRAFTKESPSMSARQVSLILFSLMPDRGKVKIESW 240  
 DB 181 LAKSPAKFVLDICIKFYLOQRRAFTKESPSMSARQVSLILFSLMPDRGKVKIESW 240

QY 241 IKDRAETAAVAWRRLMTEGGLAAAEKMDARGLLLVACFGVSNFRSTDLLDLIRMSG 300  
 DB 241 IKDRAETAAVAWRRLMTEGGLAAAEKMDARGLLLVACFGVSNFRSTDLLDLIRMSG 300

QY 301 NEIAGALKRSQFLVPMVSGIIVESSIKRGMIHEALEMYITFGMEDKFSAAVLTSFLKMSK 360  
 DB 301 NEIAGALKRSQFLVPMVSGIIVESSIKRGMIHEALEMYITFGMEDKFSAAVLTSFLKMSK 360

QY 361 ESFPRAKGAOSPLAFKEAATKQALVSSVMQCMETHKLDPAKELPWQIKQIVSLEKD 420  
 DB 361 ESFPRAKGAOSPLAFKEAATKQALVSSVMQCMETHKLDPAKELPWQIKQIVSLEKD 420

QY 421 TLQDKEMEEXKARSLSLMEEAALAKRMVNOQIKRPLSPMEMPPTVSSSYPIYDRDSP 480  
 DB 421 TLQDKEMEEXKARSLSLMEEAALAKRMVNOQIKRPLSPMEMPPTVSSSYPIYDRDSP 480

QY 481 SQRDDDDQDEIISALVSSYLGSTSPHRSRSPSPVWPLPHGGGRSVYAYEHLAPNSYSP 540  
 DB 481 SQRDDDDQDEIISALVSSYLGSTSPHRSRSPSPVWPLPHGGGRSVYAYEHLAPNSYSP 540

QY 541 GGHGHLRHQYSPLVHGQRHPLQYSPPIHGQQQLPYGIQRYVHRHSPSEERYLGLSNQSP 600  
 DB 541 GGHGHLRHQYSPLVHGQRHPLQYSPPIHGQQQLPYGIQRYVHRHSPSEERYLGLSNQSP 600

QY 601 RSNSSLDPK 609  
 DB 601 RSNSSLDPK 609

RESULT 2  
 ID AAG06057 standard; Protein; 558 AA.  
 XX  
 AC AAG06057;  
 XX  
 DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 2691.  
 DE  
 XX Protein identification; signal transduction pathway; metabolic pathway;  
 XX hybridisation assay; genetic mapping; gene expression control; promoter;  
 XX termination sequence.  
 OS Arabidopsis thaliana.  
 PN EP1033405-A2.  
 XX  
 XX 06-SEP-2000.  
 PD  
 XX 25-FEB-2000; 2000EP-0301439.  
 PF  
 XX 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 23-MAR-1999; 99US-0128264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 23-APR-1999; 99US-0130891.  
 PR 28-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 04-MAY-1999; 99US-0132407.  
 PR 05-MAY-1999; 99US-0132484.  
 PR 06-MAY-1999; 99US-0132485.  
 PR 07-MAY-1999; 99US-0132487.  
 PR 11-MAY-1999; 99US-0132863.  
 PR 14-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 17-JUN-1999; 99US-0139453.  
 PR 18-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139763.

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

DM protein - protein search, using sw model

run on: August 14, 2003, 10:27:44 ; Search time 85 Seconds  
(without alignments)  
1137.229 Million cell updates/sec

Title: US-09-890-475-1

Perfect score: 3104

Sequence: 1 MSNPPVAAQPTTANPLL.....RYLGLNQSPRSNSLDPK 609

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_19Jun03.\*  
1: /SID1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SID1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3104	100.0	609	21	AA08030
2	347	11.2	558	21	AA06057
3	346	11.1	548	21	AA06058
4	338	10.9	560	21	AA039607
5	337	10.9	550	21	AA039608
6	336	10.8	454	21	AA06059
7	327	10.5	456	21	AA039609
8	296.5	9.6	532	21	AA011239
9	296.5	9.6	532	21	AA049607

10	292.5	9.4	505	21	AA011240	Arabidopsis thalia
11	292.5	9.4	505	21	AA049608	Arabidopsis thalia
12	282	9.1	532	21	AA019753	Arabidopsis thalia
13	281.5	9.1	483	21	AA011241	Arabidopsis thalia
14	281.5	9.1	483	21	AA049609	Arabidopsis thalia
15	277	8.9	505	21	AA019754	Arabidopsis thalia
16	275	8.9	488	21	AA019755	Arabidopsis thalia
17	151.5	4.9	2421	22	AB065643	Drosophila melanog
18	146	4.7	709	23	AB070293	Human novel polype
19	144	4.6	709	23	AB070293	Human novel polype
20	144	4.6	725	23	AAU11436	Human hyaluronic a
21	143.5	4.6	1201	20	AA090345	Drosophila sp. Cos
22	143	4.6	972	18	ABJ19394	Human intracellular
23	142	4.6	725	18	AA039165	Human RHAMM protei
24	142	4.6	725	23	AB060842	Human receptor for
25	141	4.5	646	23	AB070292	Human novel polype
26	140.5	4.5	710	22	AA078497	Human protein seq
27	137.5	4.4	2074	21	AA054319	Amino acid sequenc
28	136.5	4.4	1201	22	AB058421	Drosophila melanog
29	136	4.4	1143	22	AB021341	Novel human diagno
30	132.5	4.3	888	23	AA093013	Human homologue of
31	130.5	4.2	864	22	AA039331	Human polypeptide
32	130.5	4.2	1084	22	AA041117	Human polypeptide
33	129	4.2	485	22	AA039971	Human polypeptide
34	129	4.2	771	22	AA069462	Human purified sec
35	127.5	4.1	786	20	AA055937	Human SULU3 protei
36	127.5	4.1	1001	20	AA055942	Human/Murine SULU3
37	127.5	4.1	1001	23	AB097326	Novel human protei
38	127.5	4.1	1005	24	AB047509	Breast cancer asso
39	127.5	4.1	1726	24	AA016416	Human nucleic acid
40	127.5	4.1	3256	21	AA050976	Human cell cycle p
41	127.5	4.1	3256	23	AB077188	Prostate adenocarc
42	127.5	4.1	3256	24	AB070489	Protein differenti
43	127	4.1	2816	22	AA068572	Human novel cytoke
44	126	4.1	951	22	AB053070	Intracellular traf
45	126	4.1	953	22	AB053071	Intracellular traf

#### ALIGNMENTS

##### RESULT 1

AA08030

ID AA08030 standard; Protein; 609 AA.

XX AA0803030;

XX 04-DEC-2000 (first entry)

DE Polypeptide which is encoded by the FRI gene of Arabidopsis.

XX H51; one locus-FRIGIDA; FRI gene; flowering time; blotting;

KW flower initiation; stem elongation; flower production.

OS Arabidopsis sp.

XX WO2000046358-A2.

PD 10-AUG-2000.

XX 25-JAN-2000; 2000WC-GB00197.

XX 05-FEB-1999; 99GB-0002660.

XX (PLAN-) PLANT BIOSCIENCE LTD.

XX Johanson U, West J, Dean C;

XX WPI; 2000-532899/48.

XX N-PSDB; AA063669.

XX New nucleic acid derived from the FRI locus of a plant, e.g.

PT Arabidopsis, encoding a polypeptide capable of specifically altering

```

295 AACGATCAAAAGCTCTCTTTGGGAGAGCTTCCAGAAATAATAGAGGAATGTTGGCAAAAC 354
328 GlyMetHisIleGluAlaLeuGluMetValThrPheGlyMetClnuAspLysPheSer 347
355 AAGCAGTAGTAAAGCTGTTTCATATGTTCTATAGTGTGGAAATGAGGAGATATTTAAC 414
348 AlaAlaLeuValLeuThrSerPheLeuLysMetSerLysGluSerPheGluArgAlaLys 367
415 CCTAGGACACTTTTAAACATTTATAGAGAGAGTGTAAGAATCAITTCATACATGAAT 474
368 ArgLysAlaGlnSerProLeuAlaPheLysGluAlaAlaThrLysGlnLeuAlaValLeu 387
475 GGATCAAAAGTGCACATCAAGGAATATTGTAGCAAAAGGGAGTACTAGGTGATCTG 534
388 SerSerValMetGlnCysMetGluThrHisLysLeuAspProAlaLysGluLeuProGly 407
535 AGATCGATCATCAATGTTTGAAGATTCATGAGGTGATCGATTCCTTCCAGGA 594
408 TrpGlnIleLysGluGlnIleValSerLeuGluLysAsp 420
595 TGGGAATCATGTAAGAGATAGCGAGCTTGGAGAAAGAA 633

RESULT 15
LOCUS BQ118418 637 bp mRNA linear EST 07-MAR-2003
DEFINITION EST603994 mixed potato tissues Solanum tuberosum cDNA clone STWEA72
3' end, mRNA sequence.
ACCESSION BQ118418
VERSION BQ118418.1 GI:20170380
KEYWORDS EST.
SOURCE Solanum tuberosum (potato)
ORGANISM Solanum tuberosum
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 637)
AUTHORS Buell,C.R., Hart,A., Baker,B., Tanksley,S., Fry,W., Smart,C.,
Rea, S., Griffiths,H., van der Hoeven,R., Tsai,J. and
Karamycheva,S.A.
TITLE Generation of a set of potato cDNA clones for microarray analyses
JOURNAL Unpublished
CONTACT Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-array@tigr.org
This clone can be obtained from the University of Arizona Genomics
Institute. Orders can be made through URL:
http://genome.arizona.edu/orders/
Seq primer: T7.
FEATURES
    source
        1..637
            /organism="Solanum tuberosum"
            /mol_type="mRNA"
            /cultivar="Kennebec or Binjite"
            /db_xref="taxon:4113"
            /clone="STWEA72"
            /tissue_type="mixed tissues"
            /lab_host="SOLR"
            /clone_lib="mixed potato tissues"
            /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
            XhoI; supplier: Combination of untreated and Phytophthora
            infestans-treated libraries of stolons, leaves, leaflets,
            axillary buds of stem explants, petioles, germinating eyes
            , tubers, or roots."
BASE COUNT 180 a 135 c 168 g 154 t
ORIGIN
Alignment Scores:
Pred. No.: 2.85e-24 Length: 637
Score: 371.00 Matches: 82
Percent Similarity: 56.04% Conservative: 34
Best Local Similarity: 39.61% Mismatches: 67
Query Match: 11.95% Indels: 24

```

```

DB: 13 Gaps: 3
US-09-890-475-1 (1-609) x BQ118418 (1-637)
Qy 111 ProMetLeuSerProProArgSerAsnValSerValGluThrThrValThrValSerGln 130
Db 21 CCTATGGAGAGTCTGGCTGGTTCTCCGCTATCCAGACTTCCACCGCCACCGCGCTGCAG 80
Qy 131 Pro---SerGlnGluIleValProGluThrSerAsnLysProGluGlyGlyArg----- 147
Db 81 CCGCAGACGGCTGATACACCCCTTCCACCGGCTCACCACCGTCCAGCTCGAATCAAAACC 140
Qy 148 -----MetCysGluLeu 151
Db 141 TCATCCGAAGAAAAAACTCAAATCTCCGATTCCAGCTCGAAACCTCTGCAAGAGC 200
Qy 152 MetCysSerLysGlyLeuArgLysTrpIleTyrAlaAsnIleSerAspGlnAlaLysLeu 171
Db 201 ATGTGGCGCAAGGACTTCGAAGTTACATAGTATCGAAACTCGCAGAGAAAAACACACTG 260
Qy 172 MetGluGluIleProSerAlaLeuLysLeuAlaLysGluProAlaLysPheValLeuAsp 191
Db 261 CGCGAAGAACTCACCAAGGCGTTGAACCTTGACCCCAACGCAGCAAAAGCTTGTACTTACG 320
Qy 192 CysIleGlyLysPheTyrLeuGlnGlyArgArgAlaPheThrLysGluSerProMetSer 211
Db 321 TGATGGAGACTTCTTCCGAAGGGCAAGCTTTTGATAGGATGCAGATGATT 380
Qy 212 SerAlaArgGlnValSerLeuLeuIleLeuGluSerPheLeuLeuMetProAspArgGly 231
Db 381 GCCACGAGGAGCATCTGCAITGGTTTGGAGTGTCTTCTTGTATG----- 428
Qy 232 LysGlyLysValLysIleGluSerTrpIleLysAspGluAlaGluThrAlaAlaValAla 251
Db 429 ---GGTTTCGATGTAATTGATGAGGGGTAAAGGAAGCTGCAAGCGACGAGTATA 485
Qy 252 TrpArgLysArgLeuMetThrGluGlyGlyLeuAlaAlaGluLysMetAspAlaArg 271
Db 486 TGGAGAAGAGGTTTCTTGTATGACGAGGTATAAGAAAGGCTAGTACATGATGATCGCGG 545
Qy 272 GlyLeuLeuLeuValAlaCysPheGlyValProSerAsnPheArgSerThrAspLeu 291
Db 546 GGATGCTATTGCTTATTGGGTGTTTGGGATTCACAAATTTTTCATATGAGGACGTT 605
Qy 292 LeuAspLeuIleArgMetSer 298
Db 606 AGGAGCTTGTATTCAGTGAGT 626

```

Search completed: August 25, 2003, 13:23:54  
Job time : 3528 secs

```

/tissue_type="flower"
/dev_stage="anthesis"
/clone_lib="tomato flower buds, anthesis, Cornell University"
/notes="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; supplier: Tanksley; Flower buds and flowers were taken from greenhouse plants (4-8 wks old, TA496). They were immediately frozen in liquid nitrogen and then size-separated while remaining frozen."
BASE COUNT      192 a  149 c  173 g  171 t
ORIGIN

Alignment Scores:
Pred. No.:      1,58e-25      Length:      685
Score:          385.00        Matches:      94
Percent Similarity: 53.56%    Conservative: 34
Best Local Similarity: 39.33% Mismatches:    79
Query Match:     12.40%      Indels:       32
DB:              9          Gaps:         6

US-09-890-475-1 (1-609) x AW738032 (1-685)
QY 86 IleGluAsnAlaIleAspSerLysLeuGluSerAsnGlyValValLeuAlaAlaArgAsn 105
Db 35 ATGGAGAACTGGCTGACTCGCGGTATCCAGACTCCAGCGCGCGG-----85
QY 106 AsnAsnPhHisGlnProMet-----LeuSerProPro---116
Db 86 -----CAGCAGCGCGCAGCGGTGATACACCCCTTACCGCGGTACACCGTCA 136
QY 117 ---ArgAsnAsnValSerValGluThrThrValThrValSerGlnProSerGlnGluIle 135
Db 137 GCCTCGAATCAAACTCATCCGAGAAAAGACTCAAAATCTCCGATTACAG-----190
QY 136 ValProGluThrSerAsnLysProGluGlyGlyArgMetCysGluLeuMetCysSerLys 155
Db 191 ---CTCGAAACC-----CTCTGCAAGCAGCATGTGGCGGAG 223
QY 156 GlyLeuArgLysTyrIleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGluGluIle 175
Db 224 GGACTTCGAAGTACATATTATTCGAAACTCCAGAGAAAACACACTCGCGCAAGACTC 283
QY 176 ProSerAlaLeuLysLeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLys 195
Db 284 ACCAAGGCATTGAACCTTGCCACCAATGAGCAAAAGCTTGTTAGCTGTGTGGAGAC 343
QY 196 PheTyrLeuGlnGlyArgAlaPheThrLysGluSerProMetSerSerAlaArgGln 215
Db 344 TTTCTTCTAAAGGGCAAGCTTTTGATAGGATGCTCGATGATGCCACAGGGAA 403
QY 216 ValSerLeuLeuIleLeuGluSerPheLeuLeuMetProAspArgGlyLysVal 235
Db 404 GCATCTGCATTGTTTGGAGTCTTTCTTGTGATA-----GGTTTCGAT 448
QY 236 LysIleLeuSerTrpIleLysAspGluAlaGluThrAlaAlaValAlaTrpArgLysArg 255
Db 449 GTRATTGTAAGGGTAAAGGAAGAGCTGCACAGCAGCAGTATATGGAGAAGAGG 508
QY 256 LeuMetThrGluGlyGlyLeuAlaAlaGluLysMetAspAlaArgGlyLeuLeu 275
Db 509 TTTGTTGATGAACAGGTATACAAAGGCTTAGTACAAATGCGATGCTCGGGATTGCTATTG 568
QY 276 LeuValAlaCysPheGlyValProSerAsnPheArgSerThrAspLeuAspLeuIle 295
Db 569 CTTATTGGGTGTTTGGGATTCCTCCCAATTTTACATGAGACATTAGGACTTGATT 628
QY 296 ArgMetSerGlySerAsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuVal 314
Db 629 CGACTGAGTAAATCATCGGAGAGATATCTACTGCCTCAGGAGATCAAAATGTGCTTATG 685

```

```

RESULT 14
AW738612      651 bp      mRNA      linear      EST 18-MAY-2001
LOCUS

```

```

DEFINITION EST340039 tomato flower buds, anthesis, Cornell University
Lycopersicon esculentum cDNA clone CTOD7N12 5', mRNA sequence.
ACCESSION AW738612
VERSION AW738612.1 GI:7647557
KEYWORDS EST.
SOURCE Lycopersicon esculentum (tomato)
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 651)
AUTHORS van der Hoeven,R.S., Bezzerides,J.L., Matern,A.L., Holt,I.E., Liang
,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Renning,C.M., Nierman
,W., Fraser,T.C., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
TITLE Generation of ESTs from tomato flower tissue, anthesis
JOURNAL Unpublished
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
FEATURES
source
1..651
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CTOD7N12"
/tissue_type="flower"
/dev_stage="anthesis"
/clone_lib="tomato flower buds, anthesis, Cornell University"
/notes="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; supplier: Tanksley; Flower buds and flowers were
taken from greenhouse plants (4-8 wks old, TA496). They
were immediately frozen in liquid nitrogen and then
size-separated while remaining frozen."
BASE COUNT      214 a   89 c   170 g   178 t
ORIGIN

Alignment Scores:
Pred. No.:      1,92e-24      Length:      651
Score:          373.00        Matches:      74
Percent Similarity: 61.03%    Conservative: 56
Best Local Similarity: 34.74% Mismatches:    81
Query Match:     12.02%      Indels:       2
DB:              9          Gaps:         1

US-09-890-475-1 (1-609) x AW738612 (1-651)
QY 208 SerProMetSerSerAlaArgGlnValSerLeuLeuIleLeuGluSerPheLeuLeuMet 227
Db 1 GCACCTTTGGGTCCGGGAAGGAGGCTAAAGTATTGATTTGGAGTCTGTTGTTTATG 60
QY 228 ProAspArgGlyLysGlyLysValLysIleGluSerTrpIleLysAspGluAlaGluThr 247
Db 61 ATTGGAACTACGATAAGGAAATCGAGCTTGAGAAATGGGTGAAGGAAGCGCGGAGCAT 120
QY 248 AlaAlaValAlaTrpArgLysArgLeuMetThrGluGlyGlyLeuAlaAlaGluLys 267
Db 121 ACAGCTTTAGCATGTTGGAGAGGATGAATCTCAAGGAGGA-----AGCCTCAAGAA 174
QY 268 MetAspAlaArgGlyLeuLeuLeuValAlaCysPheGlyValProSerAsnPheArg 287
Db 175 ATCGATGCCCGAGGTTTGCTCTCTTTATTGGGGTITTTGGATTCCAGATCAATTACA 234
QY 288 SerThrAspLeuAspLeuIleArgMetSerGlySerAsnGluIleAlaGlyValLeu 307
Db 235 AATGCGAATATCAGAAATTTGTTTCAGGTAAGTAACTTAAGTCGAGTGTGATGCTACT 294
QY 308 LysArgSerGlnPheLeuValProMetValSerGlyIleValGluSerSerIleLysArg 327

```



```

Pred. No.: 1,45e-30 Length: 771
Score: 440.00 Matches: 99
Percent Similarity: 56.35% Conservative: 43
Best Local Similarity: 39.29% Mismatches: 87
Query Match: 14.18% Indels: 24
DB: 12 Gaps: 3

US-09-890-475-1 (1-609) x B1176101 (1-771)
Qy 111 ProMetLeuSerProProArgAsnAsnValSerValGluThrThrValThrValSerGln 130
Db 36 CCCAGACTCCACCGCCACCGCCGCGCAG-ACGGCTGATACCACC----- 79
Qy 131 ProSerGlnGluLeuValProGluThrSerAsnLysProGluGlyGlyArg----- 147
Db 80 CTTTCACCGCGCTCACCACTGAGCTGATCAACCTCATCCGAGAAAGAACTCA 139
Qy 148 -----MetCysGluLeuMetCysSerLysGlyLeuArg 158
Db 140 AAATCTCCGATTACAGCTCGAAACCTCTGCAAGACGATGCGGCAAGGAGCTTCCA 199
Qy 159 LysPyrIleThrAlaAsnIleSerAspGlnAlaLysLeuMetGluGluIleProSerAla 178
Db 200 AGTTACATAGTATCGAACTCCAGAGAAACACACACTGCGGAGAGACTCACCAAGCA 259
Qy 179 LeuLysLeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPheTyrLeu 198
Db 260 TTGGAACCTTCACCAATGATGAGCAAGCTGTACTTAGCTGTATGGAGACTTCTTTCG 319
Qy 199 GlnGlyArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeu 218
Db 320 AAAGGGCAAGAGCTTTGATAGATGACAGATGATTCACCAAGGAGGAGCACTTCCA 379
Qy 219 LeuIleLeuGluSerPheLeuLeuMetProAspArgGlyLysGlyLysValLysIleGlu 238
Db 380 TTGGTTTGGAGTGTCTTCTTGTGTATG-----GGTTTCATGATGATTCAT 424
Qy 239 SerTrpIleLysAspGluAlaGluThrAlaAlaValAlaIleTrpArgLysArgLeuMetThr 258
Db 425 GAAGGGGTAAAGGAAGCTGCACAGCAGCAGTAAATATGAGAGAGAGGTTTGTGAT 484
Qy 259 GluGlyLysLeuAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuValAla 278
Db 485 GAACGAGTATAGAAAGCTGTACATGATGATGCGGGGATTCCTATTGCTTATGGG 544
Qy 279 CysPheGlyValProSerAsnPheArgSerThrAspLeuAspLeuIleArgMetSer 298
Db 545 TGTTTGGGATTCACAAATTTTTCATAGGAGCTTAGGAGCTTGATTCGAGTGAGT 604
Qy 299 GlySerAsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetValSer 318
Db 605 AATATCAGGAGATTTCTACTGCACTCAGGAGATCAATGTGCTAATGAAAGATTCCA 664
Qy 319 GlyIleValGluSerSerIleLysArgGlyMetHis-IleGluAlaLeuGluMetValTy 338
Db 665 GAAATATAGAGGCGATGTTGTAAGCATAAGATGGAAGTTGATGCTGTAATATGTCCTA 724
Qy 338 rThrPheGlyMetGluAspLysPheSerAlaAla 349
Db 725 TACTTTTGGAGTTAGGAGAAATGATAGCCCTTGC 758

RESULT 11
BH984132 688 bp DNA linear GSS 07-OCT-2002
LOCUS od42a06.b1 B.oleracea002 Brassica oleracea genomic, genomic survey
DEFINITION sequence.
ACCESSION BH984132
VERSION BH984132.1 GI:23516303
KEYWORDS Brassica oleracea
SOURCE Brassica oleracea
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

```

```

; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 688)
Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T., Nash
W., Rabinowicz,P.D. and Wilson,R.K.
Whole genome shotgun reads from Brassica oleracea
Unpublished
Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Plate: od42 row: a column: 06
Seq primer: -21UPOT forward
Class: shotgun
High quality sequence start: 103
High quality sequence stop: 551.
Location/Qualifiers
1..688
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_lib="B.oleracea002"
/note="Vector: pOTw13; Whole genome shotgun library from
flowering buds. DNA was purified from a crude nuclear
prep using Brassica oleracea TOL000DH3 buds provided by
Thomas Osborn at the University of Wisconsin. Genomic
DNA was provided by Pablo Rabinowicz (CSHL) and the
shotgun library prepared at Washington University Genome
Sequencing Center."
BASE COUNT 179 a 192 c 155 g 160 t 2 others
ORIGIN
Alignment Scores: 3,22e-30 Length: 688
Pred. No.: 435.50 Matches: 102
Score: 63.30% Conservative: 36
Percent Similarity: 46.73% Mismatches: 46
Best Local Similarity: 14.03% Indels: 34
Query Match: 28 Gaps: 5
DB:
US-09-890-475-1 (1-609) x BH984132 (1-688)
Qy 9 AlaAlaGlnProThrThrThrAlaAsn-----ProLeuLeuGlnArgHisGln 24
Db 104 GCCCGTGATCCATCCACGAGGAGAAATAAACCATCATCGCCGACCATACAACGG----- 157
Qy 25 SerGluGlnArgArgArgGluLeuProLysIleValGluThrGluSerThrSerMetAsp 44
Db 158 -----GGAACCGTGCTTACAAACACCGAA 181
Qy 45 IleThrIleGlyGlnSerLysGlnProGlnPheLeuLysSerIleAspGluLeuAlaAla 64
Db 182 ATCAGATCGAACATCTAACCATCGCAATTTTGAATCGATCGACGATTTAACTGCG 241
Qy 65 PheSerValAlaValGluThrPheLysArgGlnPheAspAspLeuGlnLysHisIleGlu 84
Db 242 TTTGCAGCTGCGATGCGCCTTCAACGCCCACTACGACGACTTGCAAAACACCATGGAT 301
Qy 85 SerIleGluAsnAlaIleAspSerLysLeuGluSerAsnGlyValValLeuAlaAlaArg 104
Db 302 TACATCAAGAAGCCCATTTGGCTCCAGTCTCAAAATTCAAAGGCATCATCGCCGAGTCTCC 361
Qy 105 AsnAsnAsnPheHisGlnProMetLeuSerProProArgAsnAsnValSerValGluThr 124
Db 362 TCCTCCCGATCGCAGTCT-----CCACGGAACGATGCTTCGCGAGAAACA 406
Qy 125 ThrValThrValSer-----GlnProSerGlnGluIleValProGlu 138
Db 407 GCCACGCGGTTGCCGCCACACATCGCGGCCAAGGAGACTTCTGAGACAGTACCGGAG 466
Qy 139 ThrSerAsnLysProGluGlyArgMetCysGluLeuMetCysSerLysGlyLeuArg 158
Db 467 ATTTCGATTAAGGTGGAG-----CGATTGTCGAGTTGATGTGCAGCAAGGCGCTGCGT 520

```

```

Db      750 GGAATTGAGGACAGATAT 767
||||:||||:||||:||||:
RESULT 9
3F050953
LOCUS      BF050953
DEFINITION EST436111 tomato developing/immature green fruit Lycopersicon
            esculentum cDNA clone cLEM21117 5' sequence similar to hypothetical
            protein [Arabidopsis thaliana] [J047115] [gb|AAC13626.1|AF058919
            F058919
            25 gene product, mRNA sequence.
ACCESSION BF050953
VERSION    BF050953.1 GI:10804849
KEYWORDS   EST.
SOURCE     Lycopersicon esculentum (tomato)
ORGANISM   Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE  1 (bases 1 to 557)
            Alcala, J., Vrebalov, J., White, R., van der Hoeven, R.S., Holt, I.E.,
            Liang, F., Hansen, T.S., Craven, M.B., Bowman, C.L., Ronning, C.M.,
            Nierman, W., Fraser, C.M., Martin, G.B., Giovannoni, J.J. and Tanksley
            S.D.
            Generation of ESTs from tomato fruit tissue, immature green
            Unpublished
            Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES             source
            1..557
            /organism="Lycopersicon esculentum"
            /mol_type="mRNA"
            /cultivar="TA496"
            /db_xref="taxon:4081"
            /clone="cLEM21117"
            /tissue_type="fruit"
            /dev_stage="immature green (5-35 days post-anthesis)"
            /lab_host="SOLR"
            /clone_lib="tomato developing/immature green fruit"
            /note="Vector: pBluescriptSKC+adapt; Site 1: EcoRI;
            Site 2: XhoI; Fruit were tagged at 5 dpa (0.5 cm) and
            harvested at 7 day intervals through 35 dpa. Equal masses
            of tissue from each stage were combined (including seeds
            and locules) prior to mRNA isolation."
BASE COUNT  157 a  76 c  164 g  160 t
ORIGIN
Alignment Scores:
Pred. No.:      2,52e-33      Length:      557
Score:          467.50      Matches:      92
Percent Similarity: 70.74%      Conservative: 41
Best Local Similarity: 48.94%      Mismatches: 50
Query Match:    15.06%      Indels: 5
DB:              10      Gaps: 3

US-09-890-475-1 (1-609) x BF050953 (1-557)
QY      171 LeuMetGluGluIleProSerAlaLeuLysLeuAlaLysGluProAlaLysPheValLeu 190
Db      3 CTGCTTGAGGAAGCCCTTAAAGCATTGAATCTCGCGCAATCGGCAAGCTTGATTG 62
QY      191 AspCysIleGlyLysPheTyrLeuGlnGlyArgAlaPheThrLysGluSerProMet 210
Db      63 GATTGTGTGGAGAGTTTATTATTCGAGGAGGAGCGGCATATGTTAAGGTTCCACCGTG 122
QY      211 SerSerAlaArgGlnValSerLeuLeuIleLeuGluSerPheLeuMetPro---Asp 229
Db      123 GTCATGAAGAGGAGGCTCTATATGTTTGTGATGCTCTTGTGTGAGGAATCGAC 182
QY      230 ArgGlyLysGlyLysValLysIleGluSerTrpIleLysAspGluAlaGluThrAlaAla 249
||||:||||:||||:||||:
Db      183 GAGGGA-----GTTGAGATTGAGAAAGAGGTCAAAAGAGAGCGGAGAGGACGCT 233
QY      250 ValAlaTrpArgLysArgLeuMetThrGluGlyGlyLeuAlaAlaAlaGluLysMetAsp 269
Db      234 TTAGCATGAGAAAGAGGTGATTCTGCTGAAGAGATTACGAAAGGTGTATGATATGAT 293
QY      270 AlaArgGlyLeuLeuLeuValAlaCysPheGlyValProSerAsnPheArgSerThr 289
Db      294 GCCCGGGTTTGTATTGTTCTATTGCTGTTTCGGGATTCAGGAGGACATTTCAGTAAATGAG 353
QY      290 AspleuLeuAspleuLeuIleArgMetSerGly---SerAsnGluIleAlaGlyAlaLeuLys 308
Db      354 GATATCAGGAGATTATTTCTGCGCAAGTCGTTCAAGAGACATATTTCTTCCCTACA 413
QY      309 ArgSerGlnPheLeuValProMetValSerGlyIleValGluSerSerIleLysArgGly 328
Db      414 AGATCAATGCTCTTCATGCGCAAGATTACAGAAATTATAGAGGGGATGCTGAACCAAG 473
QY      329 MetHisIleGluAlaLeuGluMetValTyrThrPheGlyMetGluAspLysPheSerAla 348
Db      474 ATGGAATGAGGACGATGATCTTCTGCTATCTTTTGGAAATGGAGGACAGATTAACTCT 533
QY      349 AlaLeuValLeuThrSerPheLeu 356
Db      534 CAGAAACTGTCGACACATATTTA 557

RESULT 10
B1176101
LOCUS      B1176101
DEFINITION EST517184 CSTs Solanum tuberosum cDNA clone cSTS24N12 5' sequence,
            mRNA sequence.
ACCESSION  B1176101
VERSION    B1176101.1 GI:14641912
KEYWORDS   EST.
SOURCE     Solanum tuberosum (potato)
ORGANISM   Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; lamids; Solanales; Solanaceae; Solanum.
REFERENCE  1 (bases 1 to 771)
            van der Hoeven, R., Bezzerides, J., Sun, H., Cho, J., Chiemiango, A.,
            Bougri, O., Buell, C.R., Ronning, C., Tanksley, S. and Baker, B.
            Generations of ESTs from sprouting potato eyes
            Unpublished
            Contact: Robin Buell
            The Institute for Genomic Research
            9712 Medical Center Dr, Rockville, MD 20850, USA
            Email: potato-array@tigr.org
            This clone can be obtained from the University of Arizona Genomics
            Institute. Orders can be made through URL:
            http://genome.arizona.edu/orders/
            Seq primer: M13P-R.
FEATURES             Location/Qualifiers
            1..771
            /organism="Solanum tuberosum"
            /mol_type="mRNA"
            /cultivar="Kennebec"
            /db_xref="taxon:4113"
            /clone="cSTS24N12"
            /tissue_type="sprouting eyes from tubers"
            /dev_stage="12-14 weeks post harvest"
            /lab_host="SOLR"
            /clone_lib="cSTS"
            /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
            XhoI; Various sizes of sprouting eyes (2mm to 15mm) were
            taken from tubers. The tubers were incubated at 26C in the
            dark for 2-3 weeks prior to sprouting. The eyes were
            frozen in liquid nitrogen immediately upon removal from
            tubers."
BASE COUNT  229 a  154 c  200 g  188 t
ORIGIN
Alignment Scores:

```

Pred. No.: 3,52e-35 Length: 520  
 Score: 487.00 Matches: 101  
 Percent Similarity: 76.87% Conservative: 12  
 Best Local Similarity: 68.71% Mismatches: 28  
 Query Match: 15.69% Indels: 6  
 DB: 29 Gaps: 3

US-09-890-475-1 (1-609) x B2512352 (1-520)

Qy 377 LysGluAlaAlaThrLysGlnLeuAlaValLeuSerSerValMetGlnCysMetGluThr 396  
 Db 91 AAAGAGGCCAACCAAAAGTTTGTAGCTGCTTGTATCATGATGAAGTGTGGAGGCT 150  
 Qy 397 HisLysLeuAspProAlaLysGluLeuProGlyTrpGlnIleLysGluGlnIleValSer 416  
 Db 151 CACAACCTTAGACCAGAGAAGTACAAGGATGCGAGTCAAGCAAGCAATGATTAAAG 210  
 Qy 417 LeuGluLysAspThrLeuGlnLeuAspLysGluMetGluLysAlaArgSerLeuSer 436  
 Db 211 TTGGAGAAGACATCTCTCACTGACAAACAGATGCAAGGGGAAGCAAGATCCATCAGT 270  
 Qy 437 LeuMetGluGluAlaAlaLeuAlaLysArgMetTyAsnGlnIleLysArgProArg 456  
 Db 271 TTAATGGAGGAACAGCATTTGACGAGAGATGTATTAACCAACAGATGAACGTCGAAG 330  
 Qy 457 LeuSerProMetGluMetProProValThrSerSerTySerProIleTyArgAsp 476  
 Db 331 TTGTCAACATGGAATGCCACAGCAGCTTCTCATCTATTCTTCTACCTACCTGCC 390  
 Qy 477 ArgSerPheProSerGlnArgAspAspGlnAspGluLysSerAlaLeuValSerSer 496  
 Db 391 CGAAGCTTCCCTAGTACACAGAC-----GATGAATATACAGCTTCTGTCAAGTAGT 441  
 Qy 497 TyrLeuGlyProSerThrSerPheProHisArgSer-----ArgArgSerProGluTyr 514  
 Db 442 TACTCTGAGCCTTACACAGGTTTCTCATCGGTCAAGTCTCAGGAGTCCCTCGATAT 501  
 Qy 515 MetValProLeuProHisGly 521  
 Db 502 TTAGTCTCA---CCTAGTGGT 519

RESULT 8  
 BQ115269 785 bp mRNA linear EST 07-MAR-2003  
 LOCUS ES7600845 mixed potato tissues Solanum tuberosum cDNA clone STMZ77  
 5', end, mRNA sequence.

ACCESSION BQ115269

VERSION BQ115269.2 GI:21916872

KEYWORDS EST

SOURCE Solanum tuberosum (potato)

ORGANISM Solanum tuberosum

REFERENCE 1 (bases 1 to 785)

AUTHORS Buell C.R., Hart A.A., Baker B., Tankalev S., Fry W., Smart C.,

Restrepo S., Griffiths H., van der Hoeven R., Tsai J. and

Karamycheva S.A.

Generation of a set of potato cDNA clones for microarray analyses

Unpublished

On Apr 17, 2002 this sequence version replaced gi:20167231.

CONTACT: Robin Buell

The Institute for Genomic Research

9712 Medical Center Dr, Rockville, MD 20850, USA

Email: potato-array@tigr.org

This clone can be obtained from the University of Arizona Genomics

Institute. Orders can be made through URL:

http://genome.arizona.edu/orders/

Seq primer: T3

Location/Qualifiers

1.785

/organism="Solanum tuberosum"

/mol\_type="mRNA"

FEATURES

source

/cultivar="Kennebec or Binjete"  
 /db\_xref="taxon:4113"  
 /clone="STMZ77"  
 /issue\_type="mixed tissues"  
 /lab\_host="SOLR"  
 /clone\_lib="mixed potato tissues"  
 /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:  
 XhoI; supplier: Combination of untreated and Phytophthora  
 infestans-treated libraries of stolons, leaves, leaflets,  
 axillary buds of stem explants, petioles, germinating eyes,  
 tubers, or roots."  
 BASE COUNT 224 a 134 c 209 g  
 ORIGIN

Alignment Scores:  
 Pred. No.: 7,08e-35 Length: 785  
 Score: 486.50 Matches: 108  
 Percent Similarity: 62.90% Conservative: 48  
 Best Local Similarity: 43.55% Mismatches: 74  
 Query Match: 15.67% Indels: 18  
 DB: 13 Gaps: 5

US-09-890-475-1 (1-509) x BQ115269 (1-785)

Qy 104 ArgAsnAsnAsnPhenHisGlnProMetLeuSerProProArgAsnAsnVal--SerValG 123  
 Db 69 AAGAACCAGAACCATCAT-----CGGAATTCATCCCTCTGAAG 107  
 Qy 123 luThrThrValThrValSerGlnProSerGlnGluIleValProGluThrSerAsnLysP 143  
 Db 108 AAGAAGAGTGGGAAGTGAATCCCTCTGTTAGAGCTA-----AAATCTACTCGTT 158  
 Qy 143 roGluGlyGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyrIleTyrA 163  
 Db 159 CAGAGCTAGAAAGCATCTCTGAAAGGACGACGCGTCTGGGAGTCTGGGAAGTACATATA 218  
 Qy 163 laAsnIleSerAspGlnAlaLysLeuMetGluGluIleProSerAlaLeuLysLeuAlaL 183  
 Db 219 CGCATCTCTGATATAAATGACTGCTTGAGGAGTCCCTAAGGCATTTGAACCTCTCGC 278  
 Qy 183 ysGluProAlaLysPheValLeuAspCysIleGlyPheTyrLeuGlnGlyArgArgA 203  
 Db 279 GCATCCGCAAGCTTGTATTGGATTGTTGGGAAGTTTATTGCAAGGAGGACAGG 338  
 Qy 203 laPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuLeuLys 223  
 Db 339 CATATGTTAAGGTTTCAACCCGTTCAATGGAAGGAGGCTTCTATATGTTGTTGGATT 398  
 Qy 223 erPheLeuLeuMetPro--AspArgGlyLysGlyLysValLysIleGluSerTrpIleL 242  
 Db 399 GCTTCTTGTGATGGGAATCGACAGGGA-----GTTGAGTTTGAGAGAGGTGA 449  
 Qy 242 ysAspGluAlaGluThrAlaAlaValAlaIleTyrArgLysArgLeuMetThrGluGlyL 262  
 Db 450 AAGAAGAGTCCGGAAGGAGGAGCTTTAGCATGGAAGAGGTTGATTGCTGAAGAGGTT 509  
 Qy 262 euAlaAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuValAlaCysPheGlyV 282  
 Db 510 TTGCAAGGCGTATGATATGATATGCCCGGGTTGCTATTGTTAATTTGGTGTTCGGA 569  
 Qy 282 alProSerAsnPheArgSerThrAspLeuLeuAspLeuIleArgMetSerGly---SerA 301  
 Db 570 TTCAGGAGAAATTTAGAAATGAGGATATCAGGAGTTTGTCTTGGCCAGTCCGTTCAAGA 629  
 Qy 301 snGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetValSerGlyIleV 321  
 Db 630 AGAATATTTCTGTTCCCTTCACAGATCAAAATGTCTTCATGGGCAAGATTACAGAAATA 689  
 Qy 321 alGluSerSerIleLysArgGlyMetHisIleGluAlaLeuGluMetValTyrThrPhe- 340  
 Db 690 TAGAGGAGATGGTGAATCAGAGATCGGAATGAGCAGTTCATATGCTATATCTTGTG 749  
 Qy 341 GlyMetGluAspLysPhe 346

Db 428 TTTCTGATCTTACCATGTTTGGCTTTTGTCTTCTTAACAGAAAGAGGCGCAACCAAG 369

Qy 383 GlnLeuAlaValLeuSerSerValMetGlnCysMetGluThrHisLysLeuSerProAla 402

Db 368 TTTTATAGTGGCTTGTATCAGTCATGAGTGTGTTGGAGGCTCACAACTTAGACCCAGAG 309

Qy 403 LysGluLeuProGlyTrpGlnIleLysGluGlnIleValSerLeuGluLysAspThrLeu 422

Db 308 AAAGAAGTCAAGGATGGCAGATCAAGAGCAATGATTAAAGTTGGAGAAGACATCTT 249

Qy 423 GlnLeuAspLysGluMetGluGluLysAlaArgSerLeuSerLeuMetGluGluAla 442

Db 248 CAACCTCGAACAACAGATCAAGAGGGAAGCAAGATCCATCAGTTTAAATGGAGGAACAGCA 189

Qy 443 LeuAlaLysArgMetTyrAsnGlnGlnIleLysArgProArgLeuSerProMetGluMet 462

Db 188 TTGACGAAGAGATTGATTAACCAACAGATGAAGCGTCCAGGTTGTGACATGGAATG 129

Qy 463 ProProValThrSerSerTyrSerProIleTyrArgAspArgSerPheProSerGln 482

Db 128 CCACGAGCAGCTTCTCATCTTATCTTCTACCTACCTGCGGAGAGCTTCCCTAGTCAC 69

Qy 483 ArgAspAspGlnAspGluLysAlaValSerSerTyrLeuGlyProSerThr 502

Db 68 AGAGAC-----GATGAATATCAGCTCTGTCTAGTAGTTACCTCGAGGCTTACCA 18

Qy 503 SerPheProHisArg 507

Db 17 GGTTCCTCATCGG 3

RESULT 6  
B77833/c  
LOCUS  
DEFINITION  
T29B23TR TAMU Arabidopsis thaliana genomic clone T29E23, genomic survey sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
B77833.1 GI:2774472

Arabidopsis thaliana (thale cress)  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

euroids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 635)  
Rounsley, S.D., Field, C.B., Bass, S., Linher, K., Golden, K.,  
Berry, K., Granger, D., Sub, E., Wible, C., Adams, M.D. and Venter, J.C.

A BAC End Sequence Database for Identifying Minimal Overlaps in Arabidopsis Genomic Sequencing. Update 3

Unpublished  
Other\_GSSs: T29E23TF

Contact: Steve Rounsley  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200  
Fax: 301 838 0208  
Email: rounsley@tigr.org

Seq primer: M13 Reverse  
Class: BAC ends  
High quality sequence stop: 635.

Location/Qualifiers  
1. .635

/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone="T29E23"  
/sex="hermaphrodite"  
/clone\_lib="TAMU"  
/note="Vector: BclOBACII; Site\_1: HindIII; Site\_2: HindIII  
; Produced by Rod Wing"

BASE COUNT 209 a 110 c 143 g 173 t

## ORIGIN

Alignment Scores:  
Pred. NO.: 2,93e-40 Length: 635  
Score: 543.00 Matches: 98  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 17.49% Indels: 0  
DB: 28 Gaps: 0

US-09-890-475-1 (1-609) x B77833 (1-635)

Qy 512 ProGluTyrMetValProLeuProHisGlyLeuGlyArgSerValTyrAlaTyrGlu 531

Db 629 CCGGATATATGTTCCACTTCACATGCTGGTGGTATAGGAAGAGTGTATGCAATGAA 570

Qy 532 HisLeuAlaProAsnSerTyrSerProGlyHisGlyHisArgLeuHisArgGlnTyrSer 551

Db 569 CATCTGCCCCCAATTCATATTTCCAGGTCAGGACATAGACTTCATCGACAGTACTCT 510

Qy 552 ProSerLeuValHisGlyGlnArgHisProLeuGlnTyrSerProIleHisGlyGln 571

Db 509 CCGTCTTTGGTTCAGGACAGACATCCACTACAGTACTCTCTCCAATTCATGGACAA 450

Qy 572 GlnGlnLeuProTyrGlyIleGlnArgValTyrArgHisSerProSerGluGluArgTyr 591

Db 449 CAACAGTTACCATATGTTATACAAAGGTTTACAGATTCACCATCTCAAGAAAGATAT 390

Qy 592 LeuGlyLeuSerAsnGlnArgSerProArgSerAsnSerSerLeuAspProLys 609

Db 389 TTGGGTTTATCCAATCAAGGTTCTCTCGCAGTAACTCATCATTAGACCCCAAA 336

## RESULT 7

BZ512352

LOCUS

DEFINITION

BOMSL50TF BO 2.3 KB Brassica oleracea genomic clone BOMSL50,

genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; euroids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 520)

Town, C.D., Van Aken, S., Utterback, T., Xoo, H. and Fraser, C.M.

Whole genome shotgun sequencing of Brassica oleracea

Unpublished

Other\_GSSs: BOMSL50TR

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TF

Class: sheared ends.

Location/Qualifiers

1. 520

/organism="Brassica oleracea"

/mol\_type="genomic DNA"

/strain="TO1000DH3"

/db\_xref="taxon:3712"

/clone="BOMSL50"

/clone\_lib="BO 2.3 KB"

/note="Vector: pHO51; Site 1: BstXI; 2-3 kb sheared

genomic DNA inserted into pHO51 using BstXI linkers"

147 a 118 c 103 g 152 t

BASE COUNT

ORIGIN

Alignment Scores:

RESULT 4  
AV541057/c  
LOCUS  
DEFINITION  
AV541057 Arabidopsis thaliana roots Columbia linear EST 07-SEP-2000  
CDNA clone RZ159D09F 3', mRNA sequence.  
AV541057  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Arabidopsis thaliana (thale cress)  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
; eurosids II; Brassicales; Brassicaceae; Arabidopsia.  
1 (bases 1 to 552)  
Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.  
A large scale analysis of cDNA in Arabidopsis thaliana: Generation  
of 12,028 non-redundant expressed sequence tags from normalized and  
size-selected cDNA libraries  
DNA Re8. 7, 175-180 (2000)  
20363093  
MEDLINE  
10907847  
COMMENT  
Contact: Erika Asamizu  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.  
Location/Qualifiers  
1. 552  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone="RZ159D09F"  
/issue\_type="roots"  
/clone\_lib="Arabidopsis thaliana roots Columbia"  
/note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2:  
XhoI"  
BASE COUNT 174 a 97 c 117 g 164 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1.68e-51 Length: 552  
Score: 663.00 Matches: 122  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 21.36% Indels: 0  
DB: 9 Gaps: 0  
US-09-890-475-1 (1-609) x AV541057 (1-552)  
QY 488 AspGluileSerAlaLeuValSerTyrLeuGlyProSerThrSerPheProHisArg 507  
DB 551 GATGAATATACAGCTCTTGAGTAGTACCTCGGCCGCTCAACATCTTTCTCATCGC 492  
QY 508 SerArgSerProGlnTyrMetValProLeuProHisGlyGlyLeuGlyArgSerVal 527  
DB 491 TCAGAAAGATCCCGGAATATATGGTTCCATTCCTCCAGTGGGTAGGAAGAAGTGT 432  
QY 528 TyrAlaTyrGluHisLeuAlaProAsnSerTyrSerProGlyHisGlyHisArgLeuHis 547  
DB 431 TATCATATGACATCTCGCCCAATTCATATTCCTCCAGGTCACGACATAGACTTCAT 372  
QY 548 ArgGlnTyrSerProSerLeuValHisGlyGlnArgHisProLeuGlnTyrSerProPro 567  
DB 371 CGACAGTACTCTCGCTCTTTGGTTCCGGACAGACATCCACTACAGTCTCTCTCCA 312  
QY 568 IleHisGlyGlnGlnLeuProTyrGlyIleGlnArgValTyrArgHisSerProSer 587  
DB 311 ATTCATGGACACACAGTATACCATGGTATACAAAGGTTTACAGACATTCACCACT 252  
QY 588 GluGluArgTyrLeuGlyLeuSerAsnGlnArgSerProArgSerAsnSerSerLeuAsp 607  
DB 251 GAAGAAAGATATTGGGTTTTTATCCAAATCAAAGGTCTCTCGCAGTAACATCATATTAG 192

QY 608 ProLys 609  
DB 191 CCCAAA 186  
RESULT 5  
BH470959/c  
LOCUS  
DEFINITION  
BH470959 BOHLJ19TR BOHL Brassica oleracea genomic clone BOHLJ19, genomic  
survey sequence.  
BH470959  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Brassica oleracea  
Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
; eurosids II; Brassicales; Brassicaceae; Brassica.  
1 (bases 1 to 750)  
Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.  
Whole genome shotgun sequencing of Brassica oleracea  
Unpublished  
Other\_GSSs: BOHLJ19TF  
Contact: Chris Town  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TR  
Class: sheared ends.  
Location/Qualifiers  
1. 750  
/organism="Brassica oleracea"  
/mol\_type="genomic DNA"  
/strain="TC1000DH3"  
/db\_xref="taxon:3712"  
/clone="BOHLJ19"  
/clone\_lib="BOHL"  
/note="Vector: pHOS1; Site\_1: BstXI; 2-3 kb sheared  
genomic DNA inserted into pHOS1 using BstXI linkers"  
BASE COUNT 222 a 160 c 150 g 218 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1.43e-46 Length: 750  
Score: 612.00 Matches: 135  
Percent Similarity: 70.22% Conservative: 23  
Best Local Similarity: 60.00% Mismatches: 34  
Query Match: 19.72% Indels: 35  
DB: 28 Gaps: 2  
US-09-890-475-1 (1-609) x BH470959 (1-750)  
QY 313 LeuValProMetValSerGlyIleValGluSerIleLysArgGlyMetHisIleGlu 332  
DB 666 ATATTCTATCGTCTGAGGTATAGTTTAACTATCAAGCGTGGAAAGCATATTGAA 607  
QY 333 AlaLeuGluMetValTyrThrPheGlyMetGluAspLysPheSerAlaAlaLeuValLeu 352  
DB 606 GCACITGGGATCATTTATACCTT-GGGATAGAGGATAGTTTTCGGCTTCTCGCTTCTA 548  
QY 353 ThrSerPheLeuLysMetSerLysGluSerPheGluArgAlaLysArgLysAlaGlnSer 372  
DB 547 ACTTTCATTCTTAGATGAGCAAGAGTCAATTTGAGAGGCAACAGAAAGCTCAAGCA 488  
QY 373 ProLeuAlaPheLys----- 377  
DB 487 CCGATAGCATTTGT-ATGGCCCTTCTTAGCTACCTTCATTGACTCTCTTTTTTTTCT 429  
QY 378 -----GluAlaAlaThrLys 382

2Y	397	HisLysLeuAspProAlaLysGluLeuProGlyTyrGlnIleLysGluGlnIleValSer	416
2Y	598	CACAAGTTAGACCCAGTGAAGAAGTACCAGGCTGGCAGATCCAAAGAGCAAATGCCGAAG	539
2Y	417	LeuGluLysAspThrLeuGlnLeuAspLysGluMetGlu	429
2Y	538	CTTGAGAAGAAGATGTTTCAGCTCGCAACAACAGATCGAAGAAGCGAGATCCATCATGTCGA	479
2Y	430	---GluLysAlaArgSerLeuMetGluGluLysAlaLeuAlaLeuAlaLysArgMetTyr	448
2Y	478	ATGGAGGAAGCGCGATCCATCAGTCTAAGGAGGAGGAGCGCAATTAGCGAGAGATTGTAT	419
2Y	449	AsnGlnGlnIleLysArgProArgLeuSerProMetGluMetProProValThrSerSer	468
2Y	418	AACCAACAGATGAACGTCGAAGTTGTCAGAAGGAAATGCCAACACAGCTTTCCTTA	359
2Y	469	SerTyrSerProIleTyrArgAspArgSerPheProSerGlnArgAspAspGlnAsp	488
2Y	358	TCTTATCTCTCTATGTACCGCGACCAAGCTTCCCTAGTCACAGAGAGGAGATGCAGAT	299
2Y	489	GluIleSerAlaLeuValSerSerTyrLeuGlyProSerThrSerPheProHisArgSer	508
2Y	298	GAATATATCAGCTCTGTGTAGTAGTACTCTCGGCCCATCAGCAGGTTTCTCTCATCGGTCA	239
2Y	509	-----ArgArgSerProGluTyrMetValProLeuProHisGlyGlyLeuGlyArgSer	526
2Y	238	AGTCTCAGAGATCCCTCGAATATATGCTT-----CCACCTGGTGGGTAGGAAGAAGT	185
2Y	527	ValTyrAlaTyrGluHisLeuAlaProAsnSerTyrSerProGlyHisGlyHisArgLeu	546
2Y	184	GTCTCTGGGTATGATCATCATCAGCTCCAAATCTTATTCTCCG-----GTT	140
2Y	547	HisArgGlnTyrSerProSerLeuValHisGlyGlnArgHisProLeuGlnTyrSerPro	566
2Y	139	TCAGAAGAGTACTCTCCA-----GTTACGAGCAGAGACTTCTCTCAAGAGTACTCTCT	86
2Y	567	ProIleHisGlyGlnGlnGlnLeuProTyrGlyIleGlnArgValTyrArgHisSerPro	586
2Y	85	CCAGTTCATGGCAACACCAAAATGCCATATGGTCTA-----TACAGACATTACCA	35
2Y	587	SerGluGluArgTyrLeuGlyLeuSer	595
2Y	34	TCGTGAATTCACACCATGGAGCTCAGC	8
RESULT	3		
LOCUS	BH479860		
DEFINITION	BOGRJ75TF BOGR Brassica oleracea genomic clone BOGRJ75, genomic survey sequence.	781 bp	DNA linear GSS 13-DEC-2001
ACCESSION	BH479860		
VERSION	BH479860.1	GI:17687964	
KEYWORDS	GSS.		
SOURCE	Brassica oleracea		
ORGANISM	Brassica oleracea		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.		
AUTHORS	1 (bases 1 to 781)		
TITLE	Town, C.D., Van Aken, S., Uterback, T., Koo, H. and Fraser, C.M.		
JOURNAL	whole genome shotgun sequencing of Brassica oleracea		
COMMENT	Unpublished		
	Other_GSSs: BOGRJ75TR		
	Contact: Chris Town		
	TIGR		
	9712 Medical Center Drive, Rockville, MD 20850, USA.		
	Tel: 301-838-3523		
	Fax: 301-838-0208		
	Email: cdtown@tigr.org		
	DNA is from a doubled haploid provided by Tom Osborn.		
	Seq primer: 1F		
	Class: sheared ends.		
FEATURES	Location/Qualifiers		
source	1..781		

REFERENCE 1 (bases 1 to 704)  
 Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.  
 Whole genome shotgun sequencing of Brassica oleracea  
 TITLE  
 JOURNAL  
 COMMENT  
 Other\_GSSs: BOHBH16TF  
 Contact: Chris Town  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA.  
 Tel: 301-838-3523  
 Fax: 301-838-0208  
 Email: cdtown@tigr.org  
 DNA is from a doubled haploid provided by Tom Osborn.  
 Seq primer: TR  
 Class: sheared ends.

FEATURES  
 source  
 1..704  
 /organism="Brassica oleracea"  
 /mol\_type="genomic DNA"  
 /strain="T01000DH3"  
 /db\_xref="taxon:3712"  
 /clone\_lib="BOHBH16"  
 /note="Vector: BOHB"  
 /note="Vector: PHOS1; Site 1: BstXI; 2-3 kb sheared  
 genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 173 a 165 c 157 g 209 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 9,72e-59 Length: 704  
 Score: 742.50 Matches: 157  
 Percent Similarity: 74.80% Conservative: 27  
 Best Local Similarity: 63.82% Mismatches: 35  
 Query Match: 23.92% Indels: 28  
 DB: 28 Gaps: 7

US-09-890-475-1 (1-609) x BH491272 (1-704)

Qy 372 SerProLeuAlaPhe-----LysGluAlaAlaThrLysGlnLeuAlaVal 386  
 Db 701 TCGGTGTTGGCTTTTATTCTTCAACAGAAACTCGAAGCAAGAAAGCAGTTGGATGCG 642  
 Qy 387 LeuSerSerValMetGlnCysMetGluThrHisLysLeuAspProAlaLysGluLeuPro 406  
 Db 641 TTATCATCAGTGATGAAGTGTGTGGAGAGCTCACAAAGTTAGACCACTGAAGAAGTACCA 582  
 Qy 407 GlyTrpGlnLysGluGlnIleValSerLeuGluLysAspThrLeuGlnLeuAspLys 426  
 Db 581 GGGTGGCAGATCCAGAGCAAAATGGCCAGCTTGAGNAGAGATTGTTGAGCTCGACAAA 522  
 Qy 427 GluMetGlu-----GluLysAlaArgSerLeuSerLeuMet 438  
 Db 521 CAGATGGAGAAGCGGAGATCCATCAGTCGAATGGAGAGCGGATCCATCAGTCTAAGG 462  
 Qy 439 GluGluAlaLeuAlaLysArgMetTyrAsnGlnGlnLysArgProArgLeuSer 458  
 Db 461 GAGGAAGCGCAATTAGCGAGAGATTGTATACCAAGATGAACGTCGAAGTTGCTCA 402  
 Qy 459 ProMetGluMetProProValThrSerSerTyrSerProIleTyrArgAspArgSer 478  
 Db 401 GAAAGGGAATGCCACCAACAGCTTCCTTATCTCTATCTCTATGACCGACCAAGC 342  
 Qy 479 PheProSerGlnArgAspAspGlnAspGlnLysSerAlaLeuValSerSerTyrLeu 498  
 Db 341 TTCCTTAGTCACAGAGGGGAGATGCGAGATGAATATCAGCTCTTGTGAGTAGTTACCTC 282  
 Qy 499 GlyProSerThrSerPheProHisArgSer-----ArgArgSerProGluTyrMetVal 516  
 Db 281 GGCCCATCAGCAGGTTTTCCTCATCGTCAAGTCTCAGGAGATCCCTGAATATATGTT 222  
 Qy 517 ProLeuProHisGlyGlyLeuGlyArgSerValTyrAlaTyrGluHisLeuAlaProAsn 536  
 Db 221 -----CCACCTGTGGGTAGSAGAAGTGTCTCTCGCTATGATCATCAGCTCCAAAT 168

Qy 537 SerTyrSerProGlyHisGlyHisArgLeuHisArgGlnTyrSerProSerLeuValHis 556  
 Db 167 TCTTATTTCTCCG-----GTTTCAAGAGGTAATCTCCA-----GTTTCAAC 129  
 Qy 557 GlyGlnArgHisProLeuGlnTyrSerProIleHisGlyGlnGlnGlnLeuProTyr 576  
 Db 128 GGACAGAGACTTCTCCAGAGTACTCTTTCAGTTCATGGGCAACCAATGCCATAT 69  
 Qy 577 GlyIleGlnArgValTyrArgHisSerProSerGluGluArgTyrLeuGlyLeuSerAsn 596  
 Db 68 GGTCTA-----TACAGACATTCACCATCTGT-GAAGATATCTTGGCTTGTCCAAAT 19  
 Qy 597 GlnArgSerProArgSer 602  
 Db 18 CACAGGACTCTCTCGTAAC 1

RESULT 2  
 BZ035667/c 706 bp DNA linear GSS 09-OCT-2002  
 LOCUS ced83e10.g1 B.oleracea002 Brassica oleracea genomic, genomic survey  
 DEFINITION sequence.  
 ACCESSION BZ035667  
 VERSION BZ035667.1 GI:23612705  
 KEYWORDS GSS.  
 SOURCE Brassica oleracea  
 ORGANISM Brassica oleracea  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
 ; eurosids II; Brassicales; Brassicaceae; Brassica.  
 REFERENCE 1 (bases 1 to 706)  
 AUTHORS Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T., Nash  
 W., Rabinowicz,P.D. and Wilson,R.K.  
 TITLE Whole genome shotgun reads from Brassica oleracea  
 JOURNAL Unpublished  
 COMMENT Contact: Richard K. Wilson  
 Genome Sequencing Center  
 Washington University School of Medicine  
 Email: submissions@watson.wustl.edu  
 Plate: ced83 row: e column: 10  
 Seq primer: -28RPpOT reverse  
 Class: shotgun  
 High quality sequence start: 17  
 High quality sequence stop: 551  
 Location/Qualifiers  
 1..706  
 /organism="Brassica oleracea"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:3712"  
 /clone\_lib="B.oleracea002"  
 /notes="Vector: pOTw13; Whole genome shotgun library from  
 flowering buds. DNA was purified from a crude nuclear  
 prep using Brassica oleracea T01000DH3 buds provided by  
 Thomas Osborn at the University of Wisconsin. Genomic  
 DNA was provided by Pablo Rabinowicz (CSHL) and the  
 shotgun library prepared at Washington University Genome  
 Sequencing Center."

BASE COUNT 177 a 164 c 159 g 204 t 2 others

ORIGIN

Alignment Scores:  
 Pred. No.: 2.05e-54 Length: 706  
 Score: 696.00 Matches: 146  
 Percent Similarity: 75.11% Conservative: 26  
 Best Local Similarity: 63.76% Mismatches: 35  
 Query Match: 22.42% Indels: 23  
 DB: 28 Gaps: 6

US-09-890-475-1 (1-609) x BZ035667 (1-706)

Qy 377 LysGluAlaAlaThrLysGlnLeuAlaValLeuSerSerValMetGlnCysMetGluThr 396  
 Db 657 AAACCTCGGAACGAAAGCAGTGATGCGTTATCATCAGTATGAGTGTGTTT-GAAGCT 599

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

3M protein - nucleic search, using frame\_plus\_p2n model

run on: August 25, 2003, 10:20:35 ; Search time 3509 Seconds  
(without alignments)  
4218.132 Million cell updates/sec

Title: US-09-890-475-1  
Perfect score: 3104  
Sequence: 1 MSNPPVAAQPTTANPLL.....RYLGSLNQSPRNSGLDPK 609

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Rgapop 6.0 , Rgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 22781392 segs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing: first 45 summaries

Command line parameters:  
-MODEL=frame\_p2n.model -DEV=xlh  
-Q/cgn2\_1/USPTO\_spoal/US09890475/runat 14082003 090350 19554/app\_query.fasta\_1.775  
-DB=EST -QFMT=fastap -SUFFIX=ret -MINMATCH=0.1 -LOOFCU=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09890475 @CGN 1 1 2810 @runat 14082003 090350 19554 -NCPU=6 -ICPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_hic.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_hic.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estom.\*  
17: em\_gss\_hum.\*  
18: em\_gss\_inv.\*  
19: em\_gss\_pln.\*  
20: em\_gss\_vrt.\*  
21: em\_gss\_fun.\*  
22: em\_gss\_mam.\*  
23: em\_gss\_mus.\*  
24: em\_gss\_pro.\*  
25: em\_gss\_rod.\*  
26: em\_gss\_phg.\*  
27: em\_gss\_vrl.\*  
28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	742.5	23.9	704	28	BH491272
C 2	696	22.4	706	28	BZ035667
C 3	687.5	22.1	781	28	BH479860
C 4	663	21.4	552	9	AV541057
C 5	612	19.7	750	28	BH470959
C 6	543	17.5	635	28	B77833
C 7	487	15.7	520	29	BZ512352
C 8	486.5	15.7	785	13	BQ115269
C 9	467.5	15.1	557	10	BF050953
C 10	440	14.2	771	12	BI176101
C 11	435.5	14.0	688	28	BH984132
C 12	423.5	13.6	671	12	BG890481
C 13	385	12.4	685	9	AW738032
C 14	373	12.0	651	9	AW738612
C 15	371	12.0	637	13	BQ118418
C 16	353.5	11.4	540	10	BE471538
C 17	352	11.3	500	13	BQ511650
C 18	321.5	10.4	2069	11	AY109393
C 19	295	9.5	544	10	BF053940
C 20	279.5	9.0	430	10	BG097426
C 21	277.5	8.9	765	14	CB970934
C 22	273	8.8	511	10	BE611371
C 23	266	8.6	421	29	CC052238
C 24	257	8.3	487	29	CNS00073
C 25	253.5	8.2	945	14	CA933488
C 26	245	7.9	595	13	BQ853859
C 27	242.5	7.8	657	14	CB342143
C 28	231.5	7.5	774	10	BG548905
C 29	231	7.4	858	10	BE661560
C 30	230.5	7.4	687	14	CA933310
C 31	229	7.4	645	14	CD480186
C 32	227.5	7.3	777	10	BQ867073
C 33	225	7.2	750	13	BUT63866
C 34	222.5	7.2	594	13	CC344181
C 35	218	7.0	800	14	CD440778
C 36	217.5	7.0	810	14	CB677353
C 37	216.5	7.0	810	14	BM779700
C 38	215.5	6.9	877	12	BQ996736
C 39	214.5	6.9	723	13	BM779770
C 40	213.5	6.9	799	12	BG593384
C 41	213	6.9	677	10	BG593384
C 42	211.5	6.8	652	12	BH464949
C 43	211.5	6.8	678	14	CB092738
C 44	210	6.8	599	13	BU042940
C 45	208.5	6.7	782	14	CB289066

ALIGNMENTS

RESULT 1  
BH491272/c  
LOCUS BOHBH16TR BOHB Brassica oleracea genomic clone BOHBH16, genomic  
DEFINITION survey sequence.  
ACCESSION BH491272  
VERSION BH491272.1 GI:17699376  
KEYWORDS GSS.  
SOURCE Brassica oleracea  
ORGANISM Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
; eurosids II; Brassicales; Brassicaceae; Brassica.

QY 135 IleValProGluThrSerAsn----- 141  
 Db 3050 CTCATGAAGACACGGCAGCTGGCCAGATCTCTCCAAACCCCAAGATCATGCCAAGGCA 3109  
 QY 142 ---LysProGluGlyGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyr 150  
 Db 3110 CCAAGAGGTGAAGAGGCAAAATCACTAAATGCGCTGCCAGTCATTACAAACAGAACCA 3169  
 QY 161 IleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGluGluIleProSerAlaLeuLys 180  
 Db 3170 ATAAACACCCCAACACACACAAACACAGTTG-----AAGGCATCC 3211  
 QY 181 LeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPhe---TyrLeuGln 199  
 Db 3212 CTGGGAAAGTAGGTGTGAAGAGAGAGCTCTAGCAGTCGGCAGAGTTCCACAGCAGCTCA 3271  
 QY 200 GlyArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeu 219  
 Db 3272 GGGGAGACACCCACACACACAGAGCCAGCAGAGATGCCAAGACATC----- 3322  
 QY 220 IleLeuGluSerPheLeuLeuMetProAsp-----ArgGlyLys 232  
 Db 3323 -----AGAACCTTTAAGGAGTCTCCAAAGCAGATCTCTGGACCCAGCAGCCCGTGTAACT 3376  
 QY 233 GlyLysValLysIleGluSerTrpIleLysAspGluAlaGluThrAlaAla---ValAla 251  
 Db 3377 GGAATGAAGATGGCCAGAACCGCTTAAGGAGAGGCGCCAGTCCTAGAGACCTGGCT 3436  
 QY 252 TrpArgLysArgLeuMetThrGluGlyGlyLeuAlaAlaGluLysMetAspAlaArg 271  
 Db 3437 GGCCTCAAGAGCTCTCCAGACACAGCTCCCTCTGAGGAATCAATGACTGATGAGAAA 3496  
 QY 272 GlyLeuLeuLeuValAlaCysPheGlyValPro----- 283  
 Db 3497 -----ACTACCAAAATAGCTTGCATAATCTCCACCAACCAAGATCAGTGGACACTCCAACA 3550  
 QY 284 -----SerAsnPheArgSerThrAspLeu-----Leu 292  
 Db 3551 AGCACAAACATGGCTTAAGAGAGTCTCAGGAAGCAGATGTAGAGGAAGATTTCTTA 3610  
 QY 293 AspLeuIleArgMetSerGlySerAsnGluIleAlaGlyAlaLeuLysArgSerGlnPhe 312  
 Db 3611 GCACCTCAGGAAACTAACACCATCA-----GCAGGG-----AAAGCCATG 3649  
 QY 313 LeuValProMetValSerGlyIleValGluSerSerIleLys-----ArgGlyMetHis 330  
 Db 3650 CTTACGCCCAACACAGCAGAGGTGATGAGAAAGACATTAAGCATTTATGGGAACCTCCA 3709  
 QY 331 IleGluAlaLeuGluMetValTyrThrPheGlyMetGluAspLysPheSerAlaAlaLeu 350  
 Db 3710 GTGCAGAAACTGGACTGGCAGGAACT----- 3736  
 QY 351 ValLeuThrSerPheLeuLysMetSerLysGluSerPheGluArgAlaLysArgLysAla 370  
 Db 3737 -----TTACCTGGCAGCAAAACAGACAGCTACAGACTCTCTAAGGAAAGGCC 3781  
 QY 371 GlnSer-----ProLeuAla---PheLysGlu----- 378  
 Db 3782 CAGGCTCTAGAGACCTGGCTGGCTTTAAAGAGCTCTTCCAGACTCTCTGGTCCACACCGAG 3841  
 QY 378 ----- 378  
 Db 3842 GAATTAGTGGCTGCTGGTAAAAACCTAAATAACCTCGGACTCTCCACAGTCAGACCCA 3901  
 QY 379 -----AlaAlaThrLysGln----- 383  
 Db 3902 GTGGACACCCCAACAGCACAAGCAACGACCCCAAGAGAGTATCAGGAAAGCAGATGTA 3961  
 QY 384 -----LeuAla-----ValLeuSerSerValMetGlnCysMetGluThr 396  
 Db 3962 GAGGGAGAACTCTTAGCGTGCAGGAATCTTAATGCCATCAGCGCAAGCCATGCACACG 4021  
 QY 397 -----His 397

Db 4022 CCTAAACCATCAGTAGGTGAAGAGAAAGACATCATATTTTGTGGAACTCCAGTGCAG 4081  
 QY 398 LysLeuAspProAlaLysGluLeuProGlyTrpGlnIleLysGluGlnIleValSerLeu 417  
 Db 4082 AAATGGACCTGACAGAAACTTAACCGGACGACAGAGCGCCCAAACTCTTAAGGAA 4141  
 QY 418 GluLysAspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSerLeuSerLeu 437  
 Db 4142 GAGGCCAGGCTCTGGAAGACCTGACTGGCTTTAAAGAGCTCTTCCAGACCCCTGGTCAT 4201  
 QY 438 MetGluGluAlaAlaLeuAlaLysArgMetTyrAsnGlnIleLysArgProArgLeu 457  
 Db 4202 ACTGAAGACAGCGGTGCTGGC-----AAAACTACTAAATG 4240  
 QY 458 SerProMetGluMetProProValThrSerSerSerTyrSerProIleTyrArgAspArg 477  
 Db 4241 CCCTGCGAATCTTCTCCACCAGATCAGACACCCCAACAAAGCAAGAGGAGCC 4300  
 QY 478 SerPheProSerGlnArgAspAspGlnAspGluIleSerAlaLeuVal-SerSerTy 497  
 Db 4301 AAGACACCTTTGGAGAAAGGAGCGTACAGAGAGCTCTCAGCCCTGAAGAGCTCA-- 4358  
 QY 497 rLeuGlyProSerThrSerPheProHisArgSerArgArgSerProGluTyrMetValPr 517  
 Db 4359 -----CACACATCAGGGGAAACCCACACACAGATATAA 4393  
 QY 517 oLeuProHisGlyGlyLeuGlyArgSerValTyrAlaTyrGluHisLeuAlaProAsnSe 537  
 Db 4394 -GTACCA--GGAGGTGAGGATAAAGCATCAACGGCTTTAGGGAAACTGCATAAACAGAA 4449  
 QY 537 rTyrSerProGlyHisGlyHisArgLeuHisArgGlnTyrSerProSerLeuValHisG1 557  
 Db 4450 ACTGAGCCCA-----GCAGCAGTGTAAGTGTAG 4479  
 QY 557 yGlnArgHisPro 561  
 Db 4480 CAAGAGGCACCCA 4492

Search completed: August 25, 2003, 15:49:25  
 Job time : 3427 secs

QY 252 TtpArglyserGluMetThrGluGlyGlyLeuAlaalaGluLysMetAspAlaArg 271  
Db 3437 GGCTTCAAGAGCTCTTCCAGACACAGGCTCCCTGAGGAATCAATGACTGATGAGAA 3496  
QY 272 GlyLeuLeuLeuValAlaCysPheGlyValPro----- 283  
Db 3497 -----ACTACCAAAATAGCTGCAATCTCCACCACCAAGATCACTGACACTCCAACA 3550  
QY 284 -----SerAsnPheArgSerThrAspLeu-----Leu 292  
Db 3551 AGCACAAGCAATGGCTTAAGAGAGTCTCAGAAAGCAGATGTAGAGGAATCTTTA 3610  
QY 293 AspLeuileArgMetSerGlySerAsnGluileAlaGlyAlaLeuLysArgSerGlnPhe 312  
Db 3611 GCACTCAGGAACCTAACACCATCA-----GCAGG-----AAAGCCATG 3649  
QY 313 LeuValProMetValSerGlyileValGluSerSerileys-----ArgGlyMetHis 330  
Db 3650 CTTACGCCCAACACAGCAGAGGTGATGAGAAAGACATTAAGCATTTATGGGAATCTCA 3709  
QY 331 IlegluAlaLeuGluMetValTyThrPheGlyMetGluAspLysPheSerAlaAlaLeu 350  
Db 3710 GTGCAGAACTGGACCTGGCAGGAACT----- 3736  
QY 351 ValLeuThrSerPheLeuLysMetSerLysGluSerPheGluArgAlaLysArgLysAla 370  
Db 3737 -----TTACCTGGCAGCAAGAAAGACAGCTACAGACTCCTTAAGGAAAGGCC 3781  
QY 371 GlnSer-----ProLeuAla-----PheLysGlu----- 378  
Db 3782 CAGGCTCTAGAGACCTGGCTGGCTTTAAGAGCTCTTCAGACTCCTGGTCACACCGAG 3841  
QY 378 ----- 378  
Db 3842 GAATTAGTGGCTGCTGTAACCAACCTTAAATAATACCTGGGACTCTCCACAGTCAGACCA 3901  
QY 379 -----AlaAlaThrLysGln----- 383  
Db 3902 GTGCACACCCCAACAGCAGCAACAGCAAGCAGCAAGAGAGTATCAGGAAGCAGATGTA 3961  
QY 384 -----LeuAla-----ValLeuSerSerValMetGlnCysMetGluThr 396  
Db 3962 GAGGAGAACTCTTAGGCTGTCAGGAATCTAATGCCATCAGCAGGCAAGCCATGTCACAG 4021  
QY 397 -----His 397  
Db 4022 CTTAAACCATCAGTAGTGAAGAGAAGACATCATATTTTGGGAAGCTCCAGTGCAG 4081  
QY 398 LysLeuAspProAlaLysGluLeuProGlyTyTrpGlnLysGluGlnLeuValSerLeu 417  
Db 4082 AAACCTGGACCTGCAGAGAACTTAACCGGCGAGCAGAGAGCGCCACAAACTCCTTAAGGAA 4141  
QY 418 GluLysAspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSerLeuSerLeu 437  
Db 4142 GAGGCCAGGCTCTGGAGACCTGAGCTTTAAGAGCTCTTCAGACCCCTGGTCTAT 4201  
QY 438 MetGluGluAlaAlaLeuAlaLysArgMetTyAsnGlnGlnLysArgProArgLeu 457  
Db 4202 ACTGAAGAAGCAGTGGCTGCTGGC-----AAAACCTACTAAATG 4240  
QY 458 SerProMetGluMetProProValThrSerSerSerTySerProIleTyArgAspArg 477  
Db 4241 CCCTGCGAATCTTCTCCACCAAGATCAGACACACCCCAACAGCAGCAGAGAGCGGCC 4300  
QY 478 SerPheProSerGlnArgAspAspGlnAspGluileSerAlaLeuVal-SerSerTy 497  
Db 4301 AAGACACCTTTGGAGAAAAGGACGTACAGAGAGCTCTCAGCCCTGAAGAAGCTCA-- 4358  
QY 497 rLeuGlyProSerThrSerPheProHisArgSerArgArgSerProGluTyMetValPr 517  
Db 4359 -----CACAGACATCAGGGGAAACCAACACACACAGATAAA 4393

QY 517 oLeuProHisGlyGlyLeuGlyArgSerValTyAlaTyArgLysLeuAlaProAsnSe 537  
Db 4394 -GTACCA--GGAGGTGAGGATATAAAGCATCAACGCGTTTAGGGAACCTGCAAAACAGAA 4449  
QY 537 rTySerProGlyHisGlyHisArgLeuHisArgGlnTySerProSerLeuValHisG 557  
Db 4450 ACTGGACCCA-----CGACCAAGTGTACTGTGTAG 4479  
QY 557 yGlnArgHisPro 561  
Db 4480 CAAGAGGCCACCCA 4492

RESULT 15  
US-09-919-039-20  
; Sequence 20, Application US/09919039  
; Publication No. US20030108871A1  
; GENERAL INFORMATION:  
; APPLICANT: Kaser, Matthew R.  
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES  
; FILE REFERENCE: PA-0035 US  
; CURRENT APPLICATION NUMBER: US/09/919,039  
; PRIOR FILING DATE: 2002-09-09  
; PRIOR APPLICATION NUMBER: 60/222,113  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 401  
; SOFTWARE: PERL Program  
; SEQ ID NO 20  
; LENGTH: 10432  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20030108871A1 2700132CB1  
US-09-919-039-20

Alignment Scores: Length: 10432  
Pred. No.: 0.0218 Matches: 138  
Score: 127.50 Conservative: 91  
Percent Similarity: 33.43% Mismatches: 230  
Best Local Similarity: 20.15% Indels: 227  
Query Match: 4.11% Gaps: 30  
DB: 11

US-09-890-475-1 (1-609) x US-09-919-039-20 (1-10432)

QY 11 GlnProThrThrThrAlaAsnProLeuLeuGlnArgHisGlnSerGluGlnArg--- 29  
Db 2717 GAGCCTTCAAAACAGTATCCACTGTAAACAGGTCAGGAAGGCTTACAGAGTTCAGGAAT 2776  
QY 30 ---ArgGluLeuProLysileValGluThrGluSerThrSerMetAspIle----- 45  
Db 2777 ATACAGAAGCTACCTGTGGAAGTGAAGTGAAGAAACAAATACAGAAATTTGTGAGTGC 2836  
QY 46 -----ThrileGlyGlnSerLysGlnProGlnPheLeuLys 57  
Db 2837 ATCCTAAAGAGAGTTCAGAGGCAACACTACTACAAAGGAGAGAGAGAGATGAAG 2896  
QY 58 SerIleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAsp 77  
Db 2897 GAAATAGAA-----AGACCTTTTTCAG 2917  
QY 78 AspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsn 97  
Db 2918 ACATATAGGAAATATTTGATTAAGAAAGAAC-----GATGAAAGATGAAGCA-- 2968  
QY 98 GlyValValLeuAlaAlaArgAsnAsnAsnPheHisGln-----ProMetLeuSer 114  
Db 2969 -----ATGAGAGATCAGAGACTTGGGGGCGAGAAATGTGCCCAATG----- 3010  
QY 115 ProProArgAsnAsnValSerValGluThrValThrValThrValSerGlnProSerGlnGlu 134  
Db 3011 -----TCTGACCTGACAGACCTCAAGAGCTTCCTGATACAGAA 3049

```
QY 333 AlaLeuGluMetValTyrThrPheGlyMetGluAspLysPheSerAlaAlaLeuValLeu 352
Db 1134325 -----ACAAATGCTCGGAATACACCGGCTGTGATCAACGAAACCTAT 1134284
QY 353 ThrSerPheLeuLysMetSerLysGluSerPheGluArgAlaLysArgLysAlaGlnSe 372
Db 1134283 AGCACTTTTGTGCAAGGTTTCA-----TGCGCTCGATGGCAAGG 1134245
QY 372 rPro--LeuAlaPheLysGluAlaAlaThrLysGlnLeuAlaValLeuSerSerValMe 391
Db 1134244 CCGGATGTTGACCATTTGGAAGGATACACCGCGATCATCTGCGATCAGGAGC----- 1134190
QY 391 rGlnCysMetGluThrHisLysLeuAspProAlaLysGluLeuProGlyTrpGlnLeuLe 411
Db 1134189 -AGATGGCGCAACCCACCGTCTA----- 1134166
QY 411 sGluGlnLeuValSerLeuGluLysAspThrLeuGlnLeuAspLysGluMetGluGluY 431
Db 1134166 ----- 1134166
QY 431 sAlaArgSerLeuSerLeuMetGluGluAlaAlaLeuAlaLysArgMetTyrAsnGlnGl 451
Db 1134166 ----- 1134166
QY 451 nLeLysArgProArgLeuSerProMetGluMetProValThrSerSerSerTyrSe 471
Db 1134165 -----CGTGGGTACCGCAACTGATGCCACCGCATGTTGGCGATTT----- 1134124
QY 471 rProileTyrArgAspArgSerPheProSerGlnArgAspAspGlnAspGluLeuLe 491
Db 1134123 -----TGTTTCCGATCCGGGAC-----CT 1134101
QY 491 rAlaLeuValSerSerTyrLeuGlyProSerThrSer-----PheProHisArgSerAr 509
Db 1134100 AACGGGGTGGCCCGGAGCTTATTCCTTCAACGTCCCTCTGTTTCCGCAT----- 1134049
QY 509 gArgSerProGluTyrMetValProLeuProHisGlyGlyLeuGlyArgSerValTyrA 529
Db 1134048 -----CGGGCCCATCATCGGTGGAAAGGCGG----- 1134021
QY 529 laTyrGluHisLeuAlaProAsnSerTyrSerProGlyHisGlyHisArgLeuHisArg 549
Db 1134020 -----AAACACCAAGCGGAGAGAGCTACCTTCAACACGCGGTGGCATGTG----- 1133973
QY 549 lnTyrSerProSerLeuValHisGlyGlnArgHisProLeuGlnTyrSerProProfil 569
Db 1133972 -----CCGAGC---GTGGAGGGCATGGGCGGGGCTCAGACATCGACCT----- 1133931
QY 569 isGlyGlnGlnLeuProTyrGlyIleGlnArgValTyrArgHisSerProSerGluG 589
Db 1133930 -----CAAAGAGCTTTTCGACGCTCCCTCTCCCTC 1133900
QY 589 luArg 590
Db 1133899 AACGA 1133895
```

## RESULT 14

```
US-09-919-172-97
; Sequence 97, Application US/0919172
; Patent No. US20020119463A1
; GENERAL INFORMATION:
; APPLICANT: Farris, Mary
; TITLE OF INVENTION: PROSTATE CANCER MARKERS
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,172
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/222,469
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PERL Program
; SEQ ID NO 97
```

```
; LENGTH: 10432
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: incyte ID No. US20020119463A1 2700132CB1
US-09-919-172-97

Alignment Scores:
Pred. No.: 0.0218 Length: 10432
Score: 127.50 Matches: 138
Percent Similarity: 33.43% Conservative: 91
Best Local Similarity: 20.15% Mismatches: 230
Query Match: 4.11% Indels: 227
DB: 10 Gaps: 30

US-09-890-475-1 (1-609) x US-09-919-172-97 (1-10432)

QY 11 GlnProThrThrAlaAsnProLeuLeuGlnArgHisGlnSerGluGlnArgArg--- 29
Db 2717 GAGCCTTCAAAACAGCATATCCACTGTAACAGAGTCTACAGAGTTCAGAGTTCAGAAAT 2776
QY 30 ---ArgGluLeuProLysIleValGluThrGluSerThrSerMetAspIle----- 45
Db 2777 ATACAGAGCTACCTGTGGAAAGTACAGAGTGAAGAAACAAATACAGAAATTTGTGATGC 2836
QY 46 -----ThrIleGlyGlnSerLysGlnProGlnPheLeuLys 57
Db 2837 ATCTTAAAGAGGTCTCAGAGGCAACACTACTACAAAGGAGAGAGAGATGAAG 2896
QY 58 SerIleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAsp 77
Db 2897 GAATAGNA-----AGACCTTTTGAG 2917
QY 78 AspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsn 97
Db 2918 ACATATAGGAAATATTTGAATTAAAGAAAGAAC-----GATGAAAGATGAAGCA--- 2968
QY 98 GlyValValLeuAlaAlaArgAsnAsnAsnHisGln-----ProMetLeuSer 114
Db 2969 -----ATGAGAGATCAAGACTTGGGGCAGAAATGTGCCACCATG----- 3010
QY 115 ProProArgAsnAsnValSerValGluThrValThrValSerGlnProSerGlnGlu 134
Db 3011 -----TCTGACCTGACAGACCTCAAGAGCTTGCCTGTATACAGAA 3049
QY 135 IleValProGluThrSerAsn----- 141
Db 3050 CTCATGAAGACACGCGCAGCTGGCCAGAACTCTCTCAAACCCCAAGATCATGCCAAGGCA 3109
QY 142 ---LysProGluGlyGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyr 160
Db 3110 CCAAAGAGTGAGAAAGCAAAATCACTAAATGCCCTGCCAGTCATTACAAACCAAGAACCA 3169
QY 161 IleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGluLeuIleProSerAlaLeuLys 180
Db 3170 ATAAACACCCCAACACACACAAACAAACAGTTG-----AAGGCATCC 3211
QY 181 LeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPhe---TyrLeuGln 199
Db 3212 CTGGGAAAGTAGGTGTCAAGAGAGCTCTAGCAGTCGGCAAGTTTCACACGACGCTCA 3271
QY 200 GlyArgArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeu 219
Db 3272 GGGGAGACCCACGACACGACAGAGAGCAGAGATGGCAAGAGCATC----- 3322
QY 220 IleLeuGluSerPheLeuLeuMetProAsp-----ArgGlyLys 232
Db 3323 -----AGAAGTTTAAAGAGTCTCAAGCAGATCTCGACCCAGCAGCCCGTGAAC 3376
QY 233 GlyLysValLysIleGluSerTrpIleLysAspGluAlaGluThrAlaAla---Valala 251
Db 3377 CGAATGAAGAGTGGCCCAAGAAACGCTTAAGGAGAGAGCCAGTCTACTAGAGACCTGGCT 3436
```



```

2000-05-17; 2000-05-17
NUMBER OF SEQ ID NOS: 159
SOFTWARE: PERL Program
SEQ ID NO 31
LENGTH: 9549
TYPE: DNA
ORGANISM: Homo sapiens
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20030124569A1 LI:350243.2:2000MAY01
FEATURE:
NAME/KEY: unsure
LOCATION: 878, 1210, 8010, 8017-8018, 8030, 8032
OTHER INFORMATION: a, t, c, g, or other
JS-10-204-887-31

Alignment Scores:
Pred. No.: 0.013 Length: 9549
Score: 129.00 Matches: 139
Percent Similarity: 30.92% Conservative: 87
Best Local Similarity: 19.02% Mismatches: 289
Query Match: 4.16% Indels: 216
JB: 14 Gaps: 27

JS-09-890-475-1 (1-609) x US-10-204-887-31 (1-9549)
2Y 2 SerAsnTyrProThrValAlaAlaGlnProThrThrThrAlaAsnProLeuLeuGln 21
Db TCCAGACAGCCACCTTCACTAGCAGCAGCCAGCCAGGCTTCTCCAGGCTTGACCGCCACA 2257
2Y 22 ArgHisGlnSerGluGlnArgArgGluLeuProLysIleValGluThr----- 38
Db GTGGCACAAGCCATCCCTCAGCTCCCACTCAAGCCCACTCAGCCCATGAGCCCACTGT 2317
3Y -----GluSerThrSerMetAspIleThrIleGlnSerLysGln 52
Db TATGGGAGAACCTTTACAGTCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2377
3Y 53 ProGlnPheLeuLysSerIle-----AspGlu 61
Db AGACAAATAAGAGAGGATCTTCAAGGACTTGTAGTCTCTGACCCCTGGGACACC 2437
3Y 62 LeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAspAspLeuGlnLys 81
Db CTTGTGCGACGAGTGTAGTCACTTATAGCCCACTTCAAGGAACTTCAGGAAATGGGATG 2497
3Y 82 HisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsnGlyValValLeu 101
Db AAAATGGAGGGCTCTCTAAATGGCTCATCACACCCCAACCAAGCCGACTGGCTAGCATC 2557
3Y 102 AlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSerProProArgAsnAsnValSer 121
Db AAGGCTGAAGCCGACAAGACTTACAGTTTTCACGGACAATGCCCCAGCCCTTCCATTGGA 2617
3Y 122 ValGluThrThrValThrValSerGlnProSerGlnGluIleValPro-----Glu 138
Db GGCAGTAGCCGCTTGAAAACACTACCTCTACTCAGCCCTGACTCTCTTACATGTGGTG 2677
3Y 139 ThrSerAsnLysProGluGlyGlyArgMetCysGluLeuMetCysSerLysGlyLeuArg 158
Db ACCAGATGGAGCTGAAGCCAGC-----TCAGTCAAAAACCAAC 2716
3Y 159 LysTyrIleTyrAlaAsnIleSerAspGlnAlaLys-----LeuMetGluGlu 174
Db AGCCCTGCTATCTCTGACATCTCTGATCTGGGAGGATGGGGAGGCGGAGGTAGACAGT 2776
3Y 175 IleProSer-----AlaLeuLysLeuAlaLysGluProAlaLysPheValLeu----- 190
Db GTCAAAATCAAGGACCGCGAACAGTGTGTAAAGAGGGGCTTAAGAAAACCTCTTTTCCC 2836
3Y -----AspCysIleGlyLysPheTyrLeuGlnGlyArgArgAlaPheThrLysGlu 207
Db CCTCAGCCTCAGAGCAAGACTCACCATATTACCAAGGCTTTGAGAGTACTATTCTCCA 2896

```

```

208 Ser-----PrometSerSerAlaArgGlnValSer 217
Db AGTTATGTCACAGTCCAGCCCTGGGGCTCTGAACCCAGCAGCAGCCAGGACGGA----- 2947
218 LeuLeuIleLeuGluSerPheLeuLeuMetProAspArgGlyLysGlyLysValLysIle 237
Db -----GTGGAGAGCCAGGCCCTGAAGACAAAGGGATGAGAACCTGAGAGCATA 2998
2948 -----
238 GluSerTrpIleLysAspGluAlaGluThrAlaAlaValAlaTrpArgLysArgLeuMet 257
Db -----
2999 GAAGGGAAGTGAAGAACCGAT----- 3019
258 ThrGluGlyGlyLeuAlaAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuVal 277
Db ----- 3019
278 AlaCysPheGlyValProSerAsnPheArgSerThrAspLeuLeuAspLeuLeuArgMet 297
Db ATCTGTGAAGAAAAGAGCCCGAGCTGAGCAGTTCAGTCAGCAGCCCTCGCTCATCCAG 3079
298 SerGlySerAsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetVal 317
Db CAGCGTCCCAATATGTACATGCGAGTCCCTGTACTACACCATGCTATGATACCCCCC 3139
318 SerGlyIleValGluSerSerIleLysArgGlyMetHisIleGluAlaLeuGluMetVal 337
Db TATGGCTACAGCAGCAGCAGAGTTACCAC-----ACCCAGCTTCTGAGCAGCTAACACGGCT 3193
338 TyrThrPheGlyMetGluAspLysPheSerAlaAlaLeuValLeuThrSerPheLeuLys 357
Db TACCGCAGCAGCAGTACGAGAA-----CAG 3217
358 MetSerLysGluSerPheGluArgAlaLysArg-----LysAlaGlnSerPro 373
Db CAGAAACGCCAGAGCTTAGAGCAGCAGCAGCGGGAGTGGCAAGAAGCAGAGATGGGC 3277
374 LeuAlaPheLysGluAlaAlaThrLysGln----- 383
Db CTGAAGGAGCGGGAGGAGCAGCTCAAGGAAAGAGTGGAAAGCCGCTCAATTCCACCA 3337
384 --LeuAlaValLeuSerSerValMetGlnCysMetGluThr----- 396
Db ACTCTCACCAGGCCCCCAGCTGACAGCAGCTGTGAATCAGGACCTGGCAAGGCCAAG 3397
397 -----HisLysLeuAspProAlaLysGlu 404
Db GAGCCAGGGGTGACCCAGCCAAATCAGTCATCATCTCCCAAGTTAGTACTCTTCANAA 3457
405 LeuProGlyTrpGlnIleLysGluGlnIleValSerLeuGluLysAspThrLeuGlnLeu 424
Db CTCCCGGCG-----CAGGCCCTGAAGCCTTAAAGTGAAGCTGAGTATGCCAGCCACTA 3514
425 AspLysGluMetGluGluLysAlaArgSerLeuSerLeuMetGluGluAlaAlaLeuAla 444
Db AGCAAGGAGGCTCTGAGGCCAAGACAGGTGCTGAGTGTGGTCGACAGGAGAGATGGAT 3574
445 LysArgMetTyrAsnGlnGlnIleLysArgProArgLeuSerProMetGluMetProPro 464
Db CCAATATCTCTGTTACCCAGCAGAGGAGAGCCCGGATGAGGACATATGTTATCTCT--- 3631
465 ValThrSerSerSerTyrSerProIle-----TyrArg 475
Db -----GCCAAGTACTCAGACATCAAGTCAGAGGATGAGCGGTGGAGGAGGAGCGG 3682
476 AspArgSerPheProSerGlnArgAspAspGlnAspGluIle----- 490
Db -----
3683 GACCGCAAAATGAAGGAGGAAGGATGCGGAGTAAGGACTCTGTCCCCCAAGGAGATGGG 3742
491 -----SerAlaLeuValSer 495
Db AAGGAAAGCACAAGTAGTACTGCAAGCTGCCACGCTCAGAGGAGTCTCCGCTTGGGAGC 3802

```



1501 GGGCTCTGAACCCACAGCCAGCCAGGA-----GTGAGAGCCAG 1542  
 225 LeuLeuMetProAspArgGlyLysGlyLysValIleLeuSerTrpIleLysAspGlu 244  
 1543 GCCCTGAAGACAAAAGGGATGAGGAACCTGAGAGCATAGAGGAAAGTGAAGAACGAT 1602  
 245 AlaGluThrAlaAlaValAlaTTPArgLysArgLeuMetThrGluGlyGlyLeuAla 264  
 1602 ----- 1602  
 265 AlaGluLysMetAspAlaArgGlyLeuLeuLeuValAlaCysPheGlyValProSer 284  
 1603 -----ATCTGTGAAGAAAAGAGGCC 1623  
 285 AsnPheArgSerThrAspLeuLeuAspLeuLeuIleArgMetSerGlySerAsnGluIleAla 304  
 1624 GAGCTGAGCAGTTCCTCACTGACGAGCCCTCGGTCAACAGCAGCGTCGCCAATATGACATG 1683  
 305 GlyAlaLeuLysArgSerGlnPheLeuValProMetValSerGlyIleValGluSerSer 324  
 1684 CAGTCCCTGTACTACACACAGTATGCTATGACCCCTATGCTACAGCGACCCAGATG 1743  
 325 IleLysArgGlyMetHisIleGluAlaLeuGluMetValThrPheGlyMetGluAsp 344  
 1744 TACCAC-----ACCCACCTTCTGAGCACTAACACGCGCTTACCGCAGCAGTACGAAGAA 1797  
 345 LysPheSerAlaAlaLeuValLeuThrSerPheLeuLysMetSerLysGluSerPheGlu 364  
 1798 -----CAGCAGAAAGCCAGAGCTTAGAG 1821  
 365 ArgAlaLysArg-----LysAlaGlnSerProLeuAlaPheLysGluAlaAla 380  
 1822 CAGCAGCAGCGGGAGTGGCAAGAGAGCAGAGATGGCTGAAGCAGCGGGAGGAGCA 1881  
 381 ThrLysGln-----LeuAlaValLeuSerSer 389  
 1882 CTCAGGAAGAGTGGAGAAAGCCGCTCAATCCACCACTCTCCAAAGGCCCCCGAC 1941  
 390 ValMetGlnCysMetGluThr----- 396  
 1942 CTGACAGACCTGGTGAATCAGGACCTGGCAAGGCCAAGGAGCGGGGTGACCCAGCC 2001  
 397 -----HSLysLeuAspProAlaLysGluLeuProGlyTrpGlnIleLys 411  
 2002 AAATCAGTCATATCCCAAGTAGATGACTCTCAAACTCCCGGCGC---CAGGCCCC 2058  
 412 GluGlnIleValSerLeuGluLysAspThrLeuGlnLeuAspLysGluMetGluGluLys 431  
 2059 GAAGCCCTTAAAGTGAAGCTGAGTGATGCCAGCCACCTAAGCAAGGAGCGCTTGAGGCC 2118  
 432 AlaArgSerLeuSerLeuMetGluGluAlaAlaLeuAlaLysArgMetTyrrAsnGlnGln 451  
 2119 AAGACAGTGTGATGTGTGTCGACGAGCAGATGGATCCAACTCTGGTACCGACAG 2178  
 452 IleLysArgProArgLeuSerProMetGluMetProProValThrSerSerTyrrSer 471  
 2179 GAGCAGAGCCCGGATGTGCACATATGTTATCT-----GCCAAGTACTCA 2226  
 472 ProIle-----TyrrArgAspArgSerPheProSerGln 482  
 2227 GACATCAAGTCAGAGATGACGCGTGAAGAGGAGCGGAGCCGCAATTTGAAGAGGAA 2286  
 483 ArgAspAspGlnAspGluIle----- 490  
 2287 AGGAGTCGAGTAAGGACTGTGTCCTCCCAAGAGATGGGAGGAAGAACAGATAGTGAC 2346  
 491 -----SerAlaLeuValSerSerTyrrLeuGlyProSerThr 502  
 2347 TGCAAGCTGCCACGTGAGAGGAGTCTCGCTGGGAGCAAGGAGCGCCGCGCAAGTGC 2406  
 503 SerPhePro-----HisArgSerArgArgSerProGluTyrrMet 515  
 2407 CATGTGCTGTGTCTCTCCCACTTACCAGCAGCAGGTC-----TAC 2448

516 ValProLeuProHisGlyGlyLeuGlyArgSerValTyrrAlaTyrrGluHisLeuAlaPro 535  
 2449 ATCCCTACATGCACGCG-----TATCTCTACAGT----- 2478  
 536 AsnSerTyrrSerProGlyHis---GlyHisArgLeuHisArgGlnTyrrSerProSerLeu 554  
 2479 CAGTCTCAGACCCCAACCCAGCTACCG-----AGCATGCTGCT 2523  
 555 ValHisGlyGlnArgHisProLeuGlnTyrrSerProProIleHisGlyGlnGlnLeu 574  
 2524 GTGATGATGCAGAACTACCCAGGTTCTTACTACCTGCGCTTCCAGCTACTCTTTTTC 2577  
 575 ProTyrrGlyIleGln----- 580  
 2578 CCATATGGCAGCAGGTCTCAGGTGGTGAAGATGCTGACAGGACGAGCCAGCCCACT 2637  
 581 ValTyrrArgHisSerProSerGluGluArgTyrrLeuGlyLeuSerAsnGlnArgSerPro 600  
 2638 GTGACTGTAAATCCAGCTCAGAGTCCAAAGCCCTGGACATCTTGCAGCAGCATGCCAGT 2697  
 601 ArgSerAsnSer 604  
 2698 CACTCAAGAGC 2709  
 RESULT 11  
 US-10-037-270-1020  
 ; Sequence 1020, Application US/10037270  
 ; Publication No. US20030104529A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tang, Y. Tom  
 ; APPLICANT: Liu, Chenghua  
 ; APPLICANT: Asundi, Vinod  
 ; APPLICANT: Zhang, Jie  
 ; APPLICANT: Ren, Feiyan  
 ; APPLICANT: Chen, Rui-hong  
 ; APPLICANT: Zhao, Qing A.  
 ; APPLICANT: Wehrman, Tom  
 ; APPLICANT: Xue, Aidong J.  
 ; APPLICANT: Yang, Yonghong  
 ; APPLICANT: Wang, Jian-Rui  
 ; APPLICANT: Zhou, Ping  
 ; APPLICANT: Ma, Yunding  
 ; APPLICANT: Wang, Dunrui  
 ; APPLICANT: Wang, Zhiwei  
 ; APPLICANT: Tillinghast, John  
 ; APPLICANT: Drmanac, Radoje T.  
 ; TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and  
 ; FILE REFERENCE: 784CIP2B  
 ; CURRENT APPLICATION NUMBER: US/10/037,270  
 ; CURRENT FILING DATE: 2002-01-04  
 ; PRIOR APPLICATION NUMBER: 09/552,317  
 ; PRIOR FILING DATE: 2000-04-25  
 ; PRIOR APPLICATION NUMBER: 09/488,725  
 ; PRIOR FILING DATE: 2000-01-21  
 ; NUMBER OF SEQ ID NOS: 1104  
 ; SOFTWARE: Pf\_Fl\_genes Version 1.0  
 ; SEQ ID NO 1020  
 ; LENGTH: 3324  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (168)..(1625)  
 ; US-10-037-270-1020  
 Alignment Scores:  
 Pred. No.: 0.00256 Length: 3324  
 Score: 129.00 Matches: 100  
 Percent Similarity: 37.95% Conservative: 81  
 Best Local Similarity: 20.96% Mismatches: 194  
 Query Match: 4.16% Indels: 102

```
QY 424 LeuAspLysGluMetGluGluLysAlaArgSerLeuMetGluGluAlaLeu 443
Db 2130 TACACTCAAGTTCGAGAGGCTACCTCTTACTAGCCC-CGTCTCGAGGCGCCTT 2188
QY 444 AlaLysArgMetTyrAsnGlnGlnLysArgProArgLeuSerProMetGluMetPro 463
Db 2189 CCCAAGGAGCTTATGC-AGCTCCCTGAGGAGACCA---TTGCTGCCATCGACGTGCTCT 2244
QY 464 ProValThrSerSerSerTyrSerProLysTyrArgAspArgSerPheProSerGln-Ar 483
Db 2245 TCGAGAACATCCGACATTCACGCGCGCCAGAGGAGGAGAGC---CCCTCCAGTGC 2301
QY 483 gAspAspGlnAspGluLysSerAlaLeuValSerSerTyrLeuGlyProSerThrSe 503
Db 2302 AGACCATGCCCGGTGTGTCTGCAGCGCTTCTCTCGTCCCATCGAGCCGCTCGGCTGCT 2361
QY 503 rPheProHisArgSerArgArgSerProGluTyrMetValProLeuProHisGlyGlyLe 523
Db 2362 ACATCCCGCGGTACCGCGGTCTCTCCCGAGCAC-----TCGCTTATGCTGGGTG 2412
QY 523 uGlyArgSerValTyrAlaTyrGluHisLeuAlaProAsnSerTyrSerProGlyHisGl 543
Db 2413 TTCCCGC-----CATGG 2424
QY 543 yHisArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGlnArgHisProLeuGl 563
Db 2425 TCGCCGCGCT-----CAACAAGATTGTGTTCG 2451
QY 563 nTyrSerProPro-----IleHisGlyGlnGlnLeuPro----- 575
Db 2452 CCTCTCTCCCGCGCGGAGCAACATCATCTCCGAGATTGTCACGCTGCTCACAGG 2511
QY 576 -TyrGly---IleGlnArgValTyrArgHisSerProSer----- 587
Db 2512 TTGGCGCGAGTCCATCGTCTTGGCGCGGTGCGCCAGCCCGTAGCTGCCATGCGCTACG 2571
QY 588 ----GluGluArgTyrLeuGlyLeuSerAsnGlnArgSerProArgSerAsnSerSerLe 606
Db 2572 GCACCGAGAGCATCACCAAGGTCGACAGATTCTCGGCC-CCGGTAACCAAGTTCGTCTACT 2630
QY 606 uAspProLys 609
Db 2631 GCTGCCAAGA 2640
```

## RESULT 10

US-10-037-270-372

Sequence 372, Application US/10037270

Publication No. US20030104529A1

## GENERAL INFORMATION:

APPLICANT: Tang, Y. Tom

APPLICANT: Liu, Chenghua

APPLICANT: Asundi, Vinod

APPLICANT: Zhang, Jie

APPLICANT: Ren, Feiyan

APPLICANT: Chen, Rui-hong

APPLICANT: Zhao, Qing A.

APPLICANT: Wehrman, Tom

APPLICANT: Xue, Aidong J.

APPLICANT: Yang, Yonghong

APPLICANT: Wang, Jian-Rui

APPLICANT: Zhou, Ping

APPLICANT: Ma, Yunqiang

APPLICANT: Wang, Dunrui

APPLICANT: Wang, Zhiwei

APPLICANT: Tillinghaast, John

APPLICANT: Drmanac, Radoje T.

TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and

FILE REFERENCE: 784CIP2B

CURRENT APPLICATION NUMBER: US/10/037,270

CURRENT FILING DATE: 2002-01-04

PRIOR APPLICATION NUMBER: 09/552,317

PRIOR FILING DATE: 2000-04-25

```

; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 372
; LENGTH: 7393
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (118)..(3012)
; US-10-037-270-372
```

```
Alignment Scores:
Pred. No.: 0.00604 Length: 7393
Score: 130.50 Matches: 137
Percent Similarity: 30.80% Conservative: 86
Best Local Similarity: 18.92% Mismatches: 274
Query Match: 4.20% Indels: 227
DB: 14 Gaps: 29
```

US-09-890-475-1 (1-609) x US-10-037-270-372 (1-7393)

```
QY 1 MetSerAsnTyrPro-----ProThrValAlaAlaGlnPro 12
Db 859 ATGCCCAACAGTCCCAACTCAAGCCCATTCAGCCCAAGCCCACTGTTATGGGAGAACCT 918
QY 13 ThrThrThrAlaAsnProLeuLeu-----GlnArgHisGlnSerGluGlnArg 28
Db 919 ---TTCAACAGTCAACCTGCTTGACTCCAGCCAGGACAGAAAGAAAGACACAAAA 975
QY 29 ArgArgGluLeuProLysIleValGluThrGluSerThrSerMetAspIleThrIleGly 48
Db 976 AAGAAGGAATCTTCAAGAGAACTTGAAGTCTCTGACCCCTCGGAAAGGTGTGTGAGCA 1035
QY 49 GlnSerLysGlnProGlnPheLeuLysSerIleAspGluLeuAlaAlaPheSerValAla 68
Db 1036 GAGGAAGCAAGCAAGCCATTC----- 1056
QY 69 ValGluThrPheLysArgGlnPheAspAspLeuGlnLysHisIleGluSerIleGluAsn 88
Db 1057 -----AGGGAATCTTCAGAAATGGGATGAAATGGAGGGGCTCTCTAAAT 1101
QY 89 AlaIleAspSerLysLeuGluSerAsnGlyValValLeuAlaAlaArgAsnAsnAsnPhe 108
Db 1102 GGCTCATCAGACCCCAACAAAGCGCTGCTGCTAGCTCAAGGCTGAAGCCGCAACATC 1161
QY 109 HisGlnProMetLeuSerProProArgAsnAsnValSerValGluThrThrVal 128
Db 1162 TACAGTTTCAGGACAAATGCCCCAGCCCTTCATTTGGAGGCGAGTAGCGGCTTGAAGAC 1221
QY 129 SerGlnProSerGlnGluLeuValPro-----GluThrSerAsnLysProGluGly 145
Db 1222 ACTACCCCTACTCAGCCCTGACTCCTTACATGTGTGTGACCCCAAGATGGAGCTGAAGCC 1281
QY 146 GlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyrIleTyrAlaAsnIle 165
Db 1282 AGC-----TCAGTCAAAACCAACAGCCCTGTCATCTCTGACATC 1320
QY 166 SerAspGlnAlaLys-----LeuMetGluGluIleProSer-----AlaLeu 179
Db 1321 TCTGATGCTGGGAGGATGGGAGGCGCAAGGTAGACAGTGTCAAATCAAGAGCGCGGAA 1380
QY 180 LysLeuAlaLysGluProAlaLysPheValLeu-----AspCysIleGly 194
Db 1381 CAGTTGTTAAAGAGGGGCTTGAAGAACTCTTTTCCCTCAGCCTCAGAGCAAGAC 1440
QY 195 LysPheTyrLeuGlnGlyArgAlaPheThrLysGluSer----- 208
Db 1441 TCACCATATTACCAAGGCTTTGAGAGTTACTATTCTCCAAGTTATGCACAGTCCAGCCCT 1500
QY 209 -----ProMetSerSerAlaArgGlnValSerLeuLeuIleLeuGluSerPhe 224
```

2048 GAGCTGCTCCGGACACAGGAGGAGCTTAAGAACTGCAGCGCAGAACGGCAGAGCCAGGAG 2107  
 470 TyrSerProIleTyrArgAspArgSerPheProSerGln-----Arg 483  
 2108 GTGCTCGGGCGACACCGGACCGGAGTTGGAGAGCAGCTGGCGGTCTCTGAGGTCGAG 2167  
 484 AspAspAspGlnaAspGluIleSerAla-LeuValSerSerTyrLeuGlyProSerThrSe 503  
 2168 GCTGATCAGGTCGGGAGCTGGAAGAACAACACTCCAGCTACAAAGACCCCTCCAGCAA 2227  
 503 rPheProHisArgSerArgSerPro 512  
 2228 CTGGCAGGAGCTGTGAAGAGGCTTCCA 2255

## RESULT 9

US-09-751-962-1  
 : Sequence 1, Application US/09751962  
 : Patent No. US20020055165A1  
 : GENERAL INFORMATION:  
 : APPLICANT: CATCHESIDE, DAVID E.  
 : TITLE OF INVENTION: REAGENTS AND METHODS FOR DIVERSIFICATION  
 : OF DNA

NUMBER OF SEQUENCES: 2  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt  
 STREET: 3100 No. US20020055165A1west Center, 90 South 7th Street  
 CITY: Minneapolis  
 STATE: MN  
 COUNTRY: USA  
 ZIP: 55402

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSeq for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/751,962  
 FILING DATE: 29-Dec-2000  
 CLASSIFICATION: <Unknown>

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/977,171  
 FILING DATE: <Unknown>  
 ATTORNEY/AGENT INFORMATION:  
 NAME: SKOOG, Mark T  
 REGISTRATION NUMBER: 40,178  
 REFERENCE/DOCKET NUMBER: 10552.13US01  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 612-332-5300  
 TELEFAX: 612-332-9081  
 TELEX: <Unknown>

## INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 9775 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: Genomic DNA

## SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-751-962-1

## Alignment Scores:

Pred. No.: 0.00821 Length: 9775  
 Score: 131.00 Matches: 128  
 Percent Similarity: 33.73% Conservative: 69  
 Best Local Similarity: 21.92% Mismatches: 189  
 Query Match: 4.22% Indels: 203  
 DB: 9 Gaps: 23

US-09-890-475-1 (1-609) x US-09-751-962-1 (1-9775)

174 GluIleProSerAlaLeuLysLeuAlaLysGluProAlaLysPheValLeuAspCysIle 193  
 1046 GAGCGGCTTTTGTCTCTCCGGCTTCGATCA-----GACTGCTTC 1084

194 GlyLysPheTyrLeuGlnGlyArgArgAlaPheThrLysGluSerProMetSerSerAla 213  
 1085 CG-AGCTGCACAGTTCTTGGAGGAGGCCAGACAGACAAGAAATATACCCCTTCTTCATC- 1142  
 214 ArgGlnValSerLeuLeuIleLeuGluSerPheLeu----- 225  
 1143 AAGCCGCTTCTCGGGCCGATCTCGAACAGTTTCATCCAGGTCGCGCGCAAGGCTAACGCC 1202  
 226 -----LeuMetProAspArgGly-----LysGlyLysValLys 236  
 1203 ATCCCATCTCTGCCATCTCACTGGCTTGACAAAGAGGAGGCGCGGAAAGGTTCGCC 1262  
 237 IleGluSer-----Trp----- 240  
 1263 ATCTCCACCATCTCTCGAGCGCTCTGGAAGTCTGACCGCTCCGATGCTGCTCCCCACC 1322  
 241 ---IleLysAspGluAlaGluThrAla-----AlaValAlaTrp 252  
 1323 GTTGTGCTTGATGAGCAGCAGCAGCTGCTCTGGGTCTGGTCTACAGCAGTCCCGAGAGTGTG 1382  
 253 ArgLysArgLeuMetThrGlnGlyLeuAlaAlaGluLysMetAspAlaArgGly 272  
 1383 AAGAGGCCCTCAGGACACAGACTGGTGTCTATCAGAGCCGGAAG-----CGCGGT 1433  
 273 LeuLeuLeuValAlaLysPheGlyValProSerAsnPheArgSerThrAspLeuLeu 292  
 1434 CTCTGTGTACAAGGTGCTACTTCCGGA-----GACACTCAG 1469  
 293 AspLeuIleArgMetSerGlySer----- 300  
 1470 GAGTGTCTCGCATCTCGTTGACTCGGATAACGATGCTCTCAAGTTTGTCTGAAGCAG 1529  
 301 -----AsnGluIleAlaGlyAlaLeu 307  
 1530 AAGGGTCGTTTCTGCCACCTCGATCAGTCCGGCTGCTTGTGTCATCAAAGGCTTCCC 1589  
 308 LysArgSerGlnPheLeuValProMetValSerGlyIleValGluSerSer----- 324  
 1590 AGCTCGAGCAGACTTTGATTTTCGAGGAACAGTCTGCCCGCGAGGGCTCTTACTACTGCC 1649  
 325 -----IleLysArgGlyMetHisIleGluAlaLeuGluMetVal 337  
 1650 CGTCTCTTCTCCGATGAGAAGTAGTCCGGGCCAAGATCATCGAGGAGGCTGAGAGGCTC 1709  
 338 TyrThr-----PheGlyMetGluAspLysPheSerAlaAla 349  
 1710 TGCACCGCTCAGACCCCGCAGGAAATCGCCTTGAGGCTGCCGATCTCTTACTTTGCT 1769  
 350 LeuValLeuThrSerPheLeuLysMetSerLysGluSerPheGluArgAla----- 366  
 1770 CTTACCAGGCGCGTTGCTGCGCGGCTTACTCTTCCGATATCGAAAGGAGCCTTGACGCC 1829  
 366 ----- 366  
 1830 AAGAGCTGGAAGTCAAGCGCAGGACTGGAGATGCTAAGGTAAGTGGGCTGAGAGGAG 1889  
 367 -----LysArgLysAlaGlnSerProLeuAlaPheLysGluAlaAlaThrLysGlnLeu 384  
 1890 GGATCAAGCTCGGGCGTCCGCTCCGCTCGCCTTCCGCTCGCCTTCCGAGGAGGCC 1949  
 385 AlaValLeuSerSerValMetGlnCys-----MetGlu 395  
 1950 GCCCAGGAGACCCCTCGAAGATCACCATGAGAGCTTTCGACGCTCCCAAGGTCTCT 2009  
 396 ThrHisLysLeuAspProAlaLysGluLeuProGlyTyrGlnIleLysGluIleVal 415  
 2010 ACCGAGAGGCTCGATGCTCTCAAGCGCTCTCGCAAAAGTCGTCGATGCCATCTAC 2069  
 416 SerLeu-----GluLysAspThrLeuGln 423  
 2070 AAGATCATTTGTCCCCATCATCAGGAGCGTCCGCAAGAACGGCGACAAAGGCTGTCTGTGCG 2129

APPLICANT: Peter VEIBY  
 APPLICANT: Gordon B. MILLS  
 APPLICANT: Robert C. EAST, Jr.  
 APPLICANT: Karen LU  
 APPLICANT: Rosemarie SCHWANDT  
 APPLICANT: Xumel ZHAO  
 APPLICANT: Karen GLATT

TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,  
 Assessment, Prevention, and Therapy of Ovarian Cancer

FILE REFERENCE: MRI-030

CURRENT APPLICATION NUMBER: US/10/097,340

PRIOR FILING DATE: 2002-03-14

PRIOR APPLICATION NUMBER: 60/276,025

PRIOR FILING DATE: 2001-03-14

PRIOR APPLICATION NUMBER: 60/325,149

PRIOR FILING DATE: 2001-09-26

PRIOR APPLICATION NUMBER: 60/276,026

PRIOR FILING DATE: 2001-03-14

PRIOR APPLICATION NUMBER: 60/324,967

PRIOR FILING DATE: 2001/09/26

PRIOR APPLICATION NUMBER: 60/311,732

PRIOR FILING DATE: 2001-08-10

PRIOR APPLICATION NUMBER: 60/325,102

PRIOR FILING DATE: 2001-09-26

PRIOR APPLICATION NUMBER: 60/323,580

PRIOR FILING DATE: 2001-09-19

NUMBER OF SEQ ID NOS: 363

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 42

LENGTH: 5142

TYPE: DNA

ORGANISM: Homo sapiens

US-10-097-340-42

# Alignment Scores:

Pred. No.: 0.00209 Length: 5142  
 Score: 132.50 Matches: 122  
 Percent Similarity: 37.37% Conservative: 91  
 Best Local Similarity: 21.40% Mismatches: 236  
 Query Match: 4.27% Indels: 122  
 DB: 14 Gaps: 25

US-09-890-475-1 (1-609) x US-10-097-340-42 (1-5142)

Qy 2 SerAsnTyrProThrValAlaAlaGlnProThrThrThr----- 15  
 Db 736 TCGAGTGGCGCGCGACACGATGCTACCCCTGAACAGCGCAACCGAGCAAGCGCT 795  
 Qy 16 AlaAsnProLeuGlnArgHisGlnSerGlnArgArgArgGluLeuProLysIle 35  
 Db 796 GGACAGCGCG-CTCCACCGGACACCTTTGAGGAACCGGAGCGCCAGTCCACCAAC- 850  
 Qy 36 ValGluThrGluSerThrSerMetAspIleThrIleGlyGlnSerIleGlnProGlnPhe 55  
 Db 851 CATGGACCTCTAGCAAAATATGACACCATGTTGGGCACTTCGAACAGCCAGCCAG 910  
 Qy 56 LeuLysSerIleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGln 75  
 Db 911 AGCCAGAACCTGAGTCTCTCAGTGGCTTTAGCGTCT------CGTCAG 955  
 Qy 76 PheAspAsp-----LeuGlnLysHisIleGluSerIleGluAsnAlaIleAsp---Ser 92  
 Db 956 ACTCAGACTGGGTCTCTTACAGACTTTTGGAGGCGCGGAGGATGTCACAGACCCAC 1015  
 Qy 93 LysLeuGluSerAsnGlyValValLeuAlaAlaArgAsnAsnAsnPheHisGlnProMet 112  
 Db 1016 ATGCTGCAGTTCAAATCAACTCCAGACCTCTTCGAGAC------CAGCAGGAG 1063  
 Qy 113 LeuSerProProArgAsnAsnValSerValGluThrThrVal-----ThrValSerGln 130  
 Db 1064 GCAGCCCCACGAGGTGGGACCATATGAAGGCCACCATCTATGGCATCTCTAGGGAG 1123  
 Qy 131 ProSerGlnGluIleValProGluThrSerAsnLysProGluGlyGlyArgMetCysGlu 150

Db 1124 GGAAGCTCAGAA-----AGTGAACCTCTGTGAGGAGGAAGCTTAGTTTGTGCTGGAG 1177  
 Qy 151 -----LeuMetCysSerLysGlyLeuArgLysTyrIleTyrAla----- 163  
 Db 1178 AAGATGAGCCTCTAGTATGGTTTCTTCTGGTTCTACTAAGCCCTGCGCAGGCGAGGCT 1237  
 Qy 164 -----AsnIleSerAsp---GlnAlaLysLeuMetGluGluIle----- 175  
 Db 1238 GAGCTTACCGAAAGTGGAGGAGCTACAGCGAAAGCTGGATGAAGAGGTGAAGAAGCGG 1297  
 Qy 176 -----ProSerAlaLeuAlaLysGluProAlaLysPheValLeuAsp 191  
 Db 1298 CAGAAGCTAGAGCCATCCCAAGTTGGCTGGACGGCGAGCTGGAGAGAAACAGAGAG 1357  
 Qy 192 CysIleGlyLysPheTyrLeuGlnGlyArgAlaPheThrLysGluSerProMetSer 211  
 Db 1358 TGCAGCCGACTGCAGGAGCTGCTGGAGAGG-----AAGGGGAGGCCAGCAG 1408  
 Qy 212 SerAlaArgGlnValSerLeuLeuIleLeuGluSerPheLeuLeuMetProAspArgGly 231  
 Db 1409 AGCAACAAGGAG-----CTCCAGAACATGAAGCGCTCTTGGACCGAGGT 1453  
 Qy 232 LysGlyLysValLysIleGluSerTrpIleLysAspGluAlaGluThrAlaAlaValAla 251  
 Db 1454 GAGAT-----TTACGACATGGGCTGGAGACCCAGCGTGTGGAG 1492  
 Qy 252 TrpArgLysArgLeuMetThrGluGlyGlyLeuAlaAlaGluLysMetAspAlaArg 271  
 Db 1493 CTGCAGAACAAAGCTGAACATGTCCAGGTCCTGAGCCTCTAAGGAGGTGTACTGAAG 1552  
 Qy 272 GlyLeuLeuLeuValAlaCysPheGlyValProSerAsnPheArgSerThrAspLeu 291  
 Db 1553 GACCTGTTA-----GAGACCGCGGAACIT 1576  
 Qy 292 LeuAspLeuIleArgMetSerGlySerAsnGluIleAlaGlyAlaLeuLys---ArgSer 310  
 Db 1577 CTGGAAGAGGTC---TTGGAGGGGAAACAGCGAGTAGAGGAGCAGCTGAGCGTGGCGAG 1633  
 Qy 311 GlnPheLeuValProMetValSerGlyIleValGluSerSerIleLysArgGlyMetHis 330  
 Db 1634 CGGAGTTGACAGCCCTGAAGGGGCTGAAAGAGGAGGTAGCTCCCTGACCAGGAG 1693  
 Qy 331 IleGluAlaLeuGluMetValTyrThrPheGlyMetGluAspLysPheSerAlaAlaLeu 350  
 Db 1694 GTGGAAACATGTCGCGCAGCAGTAC----- 1717  
 Qy 351 ValLeuThrSerPheLeuLysMetSerLysGluSerPheGluArgAlaLysArgLysAla 370  
 Db 1718 -----CAGCGAGACAGAGCAGCTCCGAGGAGCATGCAAGATGCA 1759  
 Qy 371 GlnSerProLeuAlaPheLysGluAlaThrLysGlnLeuAlaValLeuSerSerVal 390  
 Db 1760 ACCCAGGACCATGCACTGCTGAGCGCGAGAGGAGGAGATGTCAGCCCTTGTGCGAGGG 1819  
 Qy 391 MetGlnCysMetGluThrHisLysLeuAspProAlaLysGluLeuProGly---TrpGln 409  
 Db 1820 CTCAG-----AGGAGCTGGAGAGACTTCAGAGAGAGCAGGCGCATTTGCAG 1867  
 Qy 410 -----IleLysGluGlnIleValSerLeu 417  
 Db 1868 AGTATGTTCCAGAGAACAGAGGAGTCTTAGAGCCACCAAGCAGGAACTCTCTGCAGCTG 1927  
 Qy 418 GluLysAspThrLeuGlnLeuAspLysGluMetGluGlyLysAlaArgSerLeuSerLeu 437  
 Db 1928 CGAATGGAGAGGAGGAGGATGGAAGAGAGCTTCGAGAGAGATAGAGGTCTTGCAGAGG 1987  
 Qy 438 MetGluGluAlaAlaLeuAlaLysArgMetTyrAsnGlnGlnIle-----LysArg 454  
 Db 1988 GAATTAGACAGCGCCCGAGCTAGTCTCGAGATACTCGCCAGGTTGAGGTGCTCAAGAAG 2047  
 Qy 455 ProArgLeuSerProMetGlu-----MetProProValThrSerSerSer 469

SEQ ID NO 1242  
LENGTH: 2526  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
IS-09-938-842A-1242

## Alignment Scores:

red. No.: 0.000481 Length: 2526  
core: 134.00 Matches: 119  
percent Similarity: 34.21% Conservative: 76  
est Local Similarity: 20.88% Mismatches: 191  
query Match: 4.32% Indels: 184  
IB: 10 Gaps: 22

IS-09-890-475-1 (1-609) x US-09-938-842A-1242 (1-2526)

Y 20 LeuGlnAArgHisGlnSerGluGlnArgArgGluLeuProLysValGluThrGlu 39  
b 394 CTTCAAAAG---CAAGTGCATCAGAACTTGAAGAAACTTAAGATTTCAGATGAAGA 450  
Y 40 SerThrSerMetAspThrIleGlyGln-----SerLysGlnProGln 54  
b 451 TACTCCAAAACAGATGCTCTTTATCTCAAGCTTTGTCCAAAACCTCTGTTCTGAGCAG 510  
Y 55 PheLeuLysSerIleAspGluLeu-----AlaAlaPheSer 66  
b 511 AACTGAATCTTTTGGAGAAATTCGAGAAAGATCTGAATTAATAATCTGCTTTGATA 570  
Y 67 ValAlaValGluThrPheLysArgGlnPheAspAspLeuGlnLysHisIleGluSerIle 86  
b 571 GTGGCTGAAGAGGAGGAAAGAGTGCAGCAATTCAGATCAGGAAATACCAAGAAAGGTA 630  
Y 87 GluAsnAlaIleAspSerLysLeuGluSerAsnGlyValValLeuAlaAlaArgAsnAsn 106  
b 631 -----AGTAAATTTGGAGTCTATCTTTGAACCAATCATCGCGCGAGAACTCA 675  
Y 107 AsnPheHisGlnProMet-----112  
b 676 GAGCTTGAAGAACTTTCAGAAATTTGCTTTCAGAAAGTGCAGAGCATCAAGATATTGGA 735  
Y 113 ---LeuSerProProArgAsn-----118  
b 736 AATGTGAGTACTAAGCGCGACGGTTGAACATAAGGTCTATTCCAAACATCTCAGTTAAA 795  
Y 119 -----AsnValSer 121  
b 796 CTAGAAAGACGAGAGAAATTTGAAGACTTGGAGCAATCAAGTGAATAACTCGAGC 855  
Y 122 ValGluThrThrValThrValSerGlnProSerGlnGluIleValProGluThrSerAsn 141  
b 856 CTTGAAGCTACTTTAAGCGTTGCAATGAGAAAGAGAGGAGCTTGTCTCAGAAATCTTAAT 915  
Y 142 LysProGluGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyrlle 161  
b 916 -----GCTGTGATGGAGAGCTGAAATCTCTTGAAGAGAGACTTGAAGAG 960  
Y 162 TyrAlaAsnIleSerAspGlnAla-----169  
b 961 CAAGCAAGAGAAATACGCAAGCTTACCACAGAGCATAGAGCTTTGAAGCTTTGCATAAA 1020  
Y 170 -----LysLeuMetGluGluIleProSerAlaLeuLysLeuAla 182  
b 1021 CACTCAGAGCTTAAAGTTCAAAGACAAATGGAGGATTTTAGTAGCAGAGATACAGAGGCC 1080  
Y 183 Lys-----GluProAlaLysPheValLeuAspCysIleGlyLysPheTyrlleGln 199  
b 1081 AAGTCITTAGCTCAGAAATCAAAAGATCTCGAGGAGAAATAAGAGTATAT-----GAA 1134  
Y 200 GlyArgArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeu 219  
b 1135 GGGAAATTTGCTGAAGCATGTGTCAATCACTGCTTTTTCGAAGAAGAACTTGATCAGTCG 1194  
Y 220 IleLeuGluSerPheLeuMetProAspArgGlyLysGlyLysValIle-----237

Db 1195 TCTGCGAGATGAGTTACTAGCAGATACAAACCAACAGCTCAAGATCAAGATTCAAGAA 1254  
Qy 238 ---GluSerTrpIleLysAspGluAlaGluThrAlaAlaValAlaTrpArgLysArgLeu 256  
Db 1255 CTTGAAGGATATCTGATCTCTGAGAGGAAC-----1287  
Qy 257 MetThrGluGlyGlyLeuAlaAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeu 276  
Db 1288 -----GCAATCGAAAAGTTA-----1302  
Qy 277 ValAlaCysPheGlyValProSerAsnPheArgSerThrAspLeuLeuAspLeuIleArg 296  
Db 1303 -----AACCAAGAGACAGAGAACTAAAGACTTGAATACA 1338  
Qy 297 MetSerGlySerAsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMet 316  
Db 1339 AAGTTCAATCCCATGAA-----AACGTAATCGAGGAGCACAAAAGCGAGCTTCTTGAA 1392  
Qy 317 ValSerGlyIleValGluSerSerIleLysArgGlyMetHisIleGluAlaLeuGluMet 336  
Db 1393 GCATCTGGAGTTGCTGATCT-----AGAAAGTGGAGGTAGAA-----1431  
Qy 337 ValTyrThrPheGlyMetGluAspLysPheSerAlaAlaLeuValLeuThrSerPheLeu 356  
Db 1432 -----GAGGCTTTGTTGAAACTTAACACTCTT 1458  
Qy 357 LysMetSerLysGluSerPheGluArg-----AlaLysArgLys 369  
Db 1459 GAGTCTCAATCAAGAGCTTGAAGAAAGATGGAGATTTGGCTGAAGTGAATATAAG 1518  
Qy 370 AlaGlnSerProLeuAlaPheLysGluAlaAlaThrLysGlnLeuAlaValLeuSerSer 389  
Db 1519 CTGAACCAAACTAGCAATCAAGGTGAGAGCCGAGATTTTCAGGCTAAGCTCTCT 1578  
Qy 390 ValMetGlnCysMetGluThrHisLysLeuAspProAlaLysGluLeuProGlyTrpGln 409  
Db 1579 GTT-----TTAGAGCTGAGAAATATCAACAGCTAAAGAACTTCAG-----1620  
Qy 410 IleLysGluGlnIleValSerLeuGluLysAspThrLeuGlnLeuAspLysGluMetGlu 429  
Db 1621 -----ATAACCATAGAGATCTACAAACAGCTTACTCTGGAAGAGAA 1665  
Qy 430 GluLysAlaArgSerLeuSerLeuMetGluGluAlaAlaLeuAlaLysArgMetTyrAsn 449  
Db 1666 AGACTACGATCAGATCTCTCCCTTGAAGAA-----GAGAAAAACCAAGTCAAT 1716  
Qy 450 GlnGlnIleLysArgProArgLeuSerProMetGluMetProProValThrSerSerSer 469  
Db 1717 -----GAGATATATCAGAGCCACCAAGATGAG 1743  
Qy 470 TyrSerProIleTyrArgAspArgSerPheProSerGlnArgAspAspGln-----487  
Db 1744 CTCGTAAGCTTCAAGCAACTCAAGTAGACAAGCTTAATCCGATGATATCGTATCC 1803  
Qy 488 -----AspGluIleSerAlaLeuValSer 495  
Db 1804 CAAATCGAGAGCTCAGTGCCTTGTGCT 1833  
RESULT 8  
US-10-097-340-42  
; Sequence 42, Application US/10097340  
; Publication No. US20030087250A1  
; GENERAL INFORMATION:  
; APPLICANT: John MONAHAN  
; APPLICANT: Manjula GANNAVAPU  
; APPLICANT: Sebastian HOERSCH  
; APPLICANT: Shubhangi KAMATKAR  
; APPLICANT: Steve G. KOVARS  
; APPLICANT: Rachel E. MEYERS  
; APPLICANT: Michael MORRISSEY  
; APPLICANT: Peter OLANDT  
; APPLICANT: Ami SEN



1210 GATCAGCTTGATTAATTACAGCAAAAGGAGGAACAAGCTGAAGAGCTGGTTCACGCAATTTG 1269  
 208 SerProMetSerSerAlaArgGlnValSerLeuLeuLeuLeuLeuLeuLeuLeuMet 227  
 : : : : :  
 1270 GAAGAGAAGCAAAAATCTAGACTCGAAGAAATTAAACCTCCCTAGAAAGAAAAGCTG----- 1323  
 228 ProAspArgGLyLysGLyLysVallYsisileGluSerTrpIleLysAspGluAlaGluThr 247  
 : : : : :  
 1324 -----AAAGGGAG-----GAGGTGAACTGGAGAAA 1350  
 248 AlaAlaValAlaTrpArgLysArgLeuMetThrGluGlyGlyLeuAlaAlaAlaGluLys 267  
 : : : : :  
 1351 AGTAGTCTGCT-----CATACCAGGCCACCTTCGTTTTGCAGGAAAAAG 1395  
 268 MetAspAlaArgGlyLeuLeuLeuLeuValAlaCySPheGlyValProSerAsnPheArg 287  
 : : : : :  
 1396 TATCACAGT-----ATGGTGCAAGCCTTGAGATGTTACTGCTCAATTGAA 1443  
 288 Ser-----ThrAspLeu 291  
 : : : : :  
 1444 AGCTATAAAGCGTTAACAGCCAGTAGAGATAGAAGATCTTAAGCTGGAGAAGCTCATATTA 1503  
 292 LeuAspLeuIleArgMetSerGlySerAsn-----GluIleAlaGly 305  
 : : : : :  
 1504 CAGGAATAAGCGCCAAGGCTGGAAAAATCGCAGGATGTTACGATCAGATTTTGGCA 1563  
 306 AlaLeuLysArgSerGlnPheLeuValProMetValSerGlyIleValGluSerSerIle 325  
 : : : : :  
 1564 ACTGNAGCTCAANTCAAGATATGTAAGATGCTTCTAGATCTCGACCAAGCTCAGCA 1623  
 326 LysArgGlyMetHisIleGluAlaLeuGluMetValTyzThrPheGlyMetGluAspLys 345  
 : : : : :  
 1624 CTAAAGGAAACAGAGAAATTAAAGAAATCACAGTTTCTTTCTCAAATAAATCACTGAT--- 1680  
 346 PheSerAlaAlaLeuValLeuThrSerPheLeuLysMetSerLysGluSerPheGluArg 365  
 : : : : :  
 1681 -----TTGCAGAACCACTTCAAGCAACAGAGGAGAGACTTTAGAAAA 1722  
 366 AlaLysArgLysAlaGlnSerProLeuAlaPheLysGluAlaAlaThrLysGlnLeuAla 385  
 : : : : :  
 1723 CAGCTGGAAGATGAAGAAGGAAGAAAGCTGAAAAAGAAAAATCAACACGAGAAATTA--- 1779  
 386 ValLeuSerSerValMetGlnCyMetGluThrHisLysLeuAspProAlaLysGluLeu 405  
 : : : : :  
 1780 -----ACTGAAGAAATT 1791  
 406 ProGlyTrpGln---IleLysGluGlnIleValSerLeuGluLysAsp---ThrLeuGln 423  
 : : : : :  
 1792 ACAAGTGGCGTCTCTCTATGAAGRACTATATAATAAAACAAAACCTTTTCAGCTACAA 1851  
 424 LeuAspLys---GluMetGluGluLysAlaArgSerLeuSerLeuMetGluGluAlaAla 442  
 : : : : :  
 1852 CTAGATGCTTTGAAGTAGAAAAAACAGGCA-----TTGTTGAATGAACATGGT 1899  
 443 LeuAlaLysArgMetTyzAsn-----Gln 450  
 : : : : :  
 1900 GCAGCTCAGGAACACCTAAATAAATAAGAGATTCATATGCTAAATATTGGTGCATCAG 1959  
 451 GlnIleLysArgProArgLeuSerProMetGluMetProProValThrSerSerSerTyz 470  
 : : : : :  
 1960 AATTTGAAACAAAAAATCAAGCATGTTGTGAAGTTGAAGATGAAATAATAGCCAATCAAA 2019  
 471 SerProIleTyzArgAspArgSerPheProSerGlnArgAspAspGlnAaspGluIle 490  
 : : : : :  
 2020 TGGGAAGTATCAAAACTCGCTGTCAGCTTGCCTTAAAAAAAACAAAGTCGACAAAACTT 2079  
 491 SerAlaLeuValSerSerTyzLeuGly-----ProSerThrSerPhePro 505  
 : : : : :  
 2080 CAAGAGGAATTGATAAAGTTCTAGGTATCAACACATTTGATCCTTCAAGGCTTTTCAT 2139  
 506 HisArgSerArgArgSerProGluTyzMetValProLeuProHisGly 521  
 : : : : :  
 2140 CATGAAAGTAAAGAAAAATTTTGCCTCGAAGACCCCCTATTAAGAGAGC 2187

```

RESULT 6
US-09-917-800A-1471
; Sequence 1471, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1471
; LENGTH: 3718
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 U67138
US-09-917-800A-1471

Alignment Scores:
Pred. No.: 0.000195 Length: 3718
Score: 140.00 Matches: 146
Percent Similarity: 35.19% Conservative: 101
Best Local Similarity: 20.80% Mismatches: 258
Query Match: 4.51% Indels: 200
DB: Gaps: 32

US-09-890-475-1 (1-609) x US-09-917-800A-1471 (1-3718)
QY 6 ProThrValAlaAlaGlnProThrThrAlaAsnProLeuLeuGlnArg----- 22
DB 1634 CCGACGGTGGCCCTCCGCGCGGAGCCGCTGCTGAAGTCCATCATACAAAGACCACCTGGA 1693
QY 23 ---HisGlnSerGlu-----GlnArgArgGluLeuProTyrIleValGluThr 38
DB 1694 GACACCAACCCAGAGTACTCTCAAGCTGCCACTGAGTGCGCT----- 1738
QY 39 GluSerThrSerMetAspIleThrIleGly-----GlnSerLysGln 52
DB 1739 GTCGGTCAACGCTGGACCCATCACTCAACTACACTTCCGAGTTCCGGTCCAGAAC 1798
QY 53 ProGlnPheLeuLysSerIleAspGluLeuAlaAlaPheSerValAlaValGluThrPhe 72
DB 1799 CAGAGCTACATCGCGGCTGTGAGCACCTTGAGCCACGCCAGCTGTGTGACCCAGATGAGT 1858
QY 73 LysArgGlnPheAspAspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSer 92
DB 1859 GAAGCG-----GAAGTTAATGGCGAGTTCGAGTCAGTGTGATCTGTCTTCAGC 1909

```

```
QY 116 ProArgAsnValSerValGluThrValThrValSerGlnProSerGlnGlu 135
DB 1495 CTCGATTCGCTTGGAGAGAGAAACCAGT---TTGAATAAAAGACAAACAAT 1551
2Y 136 ValProGluThrSerAsnLysProGluGlyGlyArgMetCysGluLeuMetCysSerLys 155
DB 1552 CAGATATGCTGAGAGAGAG---GGGACACAAGCTGGAGATACATACAGCTCAAG 1605
2Y 156 GlyLeuArgLysTyrlleTyrlleAlaAsnLysSerAspGlnAlaLysLeuMetGluGlu 175
DB 1606 CACATGTTGGATGTAAGAGCGAGGTAATGTTCTTCAGAGAAGATTGAAATCTT 1665
QY 176 ProSerAlaLeuLysLeuAlaLysGluProAlaLysPheValLeuAspCysLysleGlyLys 195
DB 1666 CAAGAGCAGCTTAGA--- 1680
QY 196 PheTyrlleGluGlnGlyArgAlaPheThrLysGluSerProMetSerSerAlaArgGln 215
DB 1681 ---GACAAGGAAAGCAGATGAGCAGCTTGAAGAA 1713
QY 216 ---ValSerLeuLeuLeuGlu 222
DB 1714 CGGTCAATCTTCAGGCTGACACCACCACTGACACTGCCTTGACAACCTTGGAG 1773
QY 223 SerPheLeuLeuMetProAspArgGlyLysValLysLysleGluSerTrpLys 242
DB 1774 GAGCCCTTCAGAGAAAGAGCGGACAATTGACGCTTAAAG---GAGCAGAGCGACAGA 1830
QY 243 AspGluAlaGluThrAlaAlaValAlaTTPArgLysArgLeuMetThrGluGlyGlyLeu 262
DB 1831 GATGAGCGAGAG--- 1842
QY 263 AlaAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuValAlaCysPheGlyVal 282
DB 1843 ---AAGCAAGAGGAAATTGAT--- 1860
QY 283 ProSerAsnPheArgSerThrAspLeuLeuAsp---LeuileArg 296
DB 1861 ---AATACAAAAA---GATCTTAAAGACTTGAAGGAAAGATCAGCGCTGTGCA 1911
QY 297 MetSerGlySerAnGluLeuAlaGlyAlaLeuLysArgSerGlnPheLeuValProMet 316
DB 1912 GCGACCTTCAGAGAAGAGGCTTCACCTTTGGAT---CTGAAGAGAGAT 1959
QY 317 ValSerGlyileValGluSerSerLysArgGlyMetHisileGluAlaLeuGluMet 336
DB 1960 GCTTCTCTCGCATCCTCAGGACTGAAAAGAGACTCAGCGCTTAAGACACTAGAGATT 2019
QY 337 ValTyrlleThrPheGlyMetGluAspLysPheSerAlaAlaLeuValLeuThrSerPheLeu 356
DB 2020 ---GCTTTGGAGAGAGAGAGGAGGAGTGTCTGAAAATGGAATCACAATTG 2067
QY 357 Lys---MetSerLysGluSerPheGluArgAla 366
DB 2068 AAAAGGCGCATGAGCGAGCATTTGGAAGCCAGAGCCAGTCCAGAGATGATGACCGAATA 2127
QY 367 LysArgLysAlaGlnSerProLeuAlaPheLysGluAlaAlaThrLysGlnLeuAlaVal 386
DB 2128 CAGCATTGGAGAGAGAGATCACAGGTACAAAGATGAATCTAGCAAGCCCGAGCAGAA 2187
QY 387 LeuSerSerValMetGlnCysMet---GluThrHisLysLeuAspProAlaLys 403
DB 2188 GTTGATCGACTCTTAGAAATCTTGAAGAGGAGTGGAAAATGAGAAGAAATGACAAGATAG 2247
QY 404 Glu---LeuProGlyTrpGlnileLysGluGln---Ile 414
DB 2248 AAGATAGCTGAGTTGGAAGATCTCAGCTCAGGCAAGTGAAGACAGAAATGAAGGTA 2307
QY 415 ValSerLeuGluLysAspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSer 434
DB 2308 GCAAATCTGAAG---CACAGGAACAGGTGGAAGAAAGAGAGAT 2349
```

```
QY 435 LeuSerLeuMetGluGluAla 441
DB 2350 GCACAAATGTTAGAGGAGGCG 2370
RESULT 5
US-09-960-253-159
; Sequence 159, Application US/09960253
; Patent No. US20020123619A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Lodes, Michael J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.556
; CURRENT APPLICATION NUMBER: US/09/960,253
; CURRENT FILING DATE: 2001-09-20
; NUMBER OF SEQ ID NOS: 187
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 159
; LENGTH: 2756
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-960-253-159
Alignment Scores:
Pred. No.: 4,52e-05 Length: 2756
Score: 144.00 Matches: 109
Percent Similarity: 35.97% Conservative: 91
Best Local Similarity: 19.60% Mismatches: 226
Query Match: 4.64% Indels: 130
DB: 20 Gaps: 20
US-09-890-475-1 (1-609) x US-09-960-253-159 (1-2756)
QY 34 LysIleValGluThrGluSerThrSerMetAspIleThrIleGlyGlnSerLysGlnPro 53
DB 706 AAATCTGTTTCAATAGAGAGAAAGAAAGATTGAT---GAAAAATCTGAAACAGAA 756
QY 54 GlnPheLeuLysSerIleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLys 73
DB 757 AAATCTTGGATACATCGAAGAAATAGTTGCTTCAGATCAAGTGGAAAAATACAG 816
QY 74 ArgGlnPheAspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSer--- 92
DB 817 CTAGATATTCGCCAGTTAGAGAAAAATTTGAAGAGAGAGAAATGATGAAATTTAAGCCTT 876
QY 93 ---LysLeuGluSerAsnGlyValValLeuAla--- 102
DB 877 AAGCAGTCTCTTGGAGACAATATTTATTTATTTATCTTAAACAAGTAGAAGATCTAAATGTG 936
QY 103 ---AlaArgAsnAsnAsnPheHisGln 110
DB 937 AAATGTCAGCTGTTGAAACAGAAAAAGAGACCATTGTCAACAGGAATAGAGAACACAAAC 996
QY 111 ProMetLeuSerProProArgAsnAsnValSerValGluThrValThrValSerGln 130
DB 997 GAAATCTTAATGACAGATGCAAAAC---TTAGACACAGAGTTTATCTTGAACAA 1050
QY 131 ProSerGlnLysLeuValProGluThrSerAsnLysProGluGlyGlyArgMetCysGln 150
DB 1051 CGGGAACATGAAAGCTTCAACAAAAAGAAATTTACAAATTCATTCTCTCGCAACAGAG 1110
QY 151 LeuMetCysSerLysGlyLeuArgLysTyrlleTyrlleAlaAsnIleSerAspGlnAlaLys 170
DB 1111 AAAGATTAATCTTCGAGTCTTCATCAGAGCTC---TGTTCT 1149
QY 171 LeuMetGluGluIleProSerAlaLeuLysLeuAlaLysGluProAlaLysPheValLeu 190
DB 1150 TTTCAAGAGGAAATGTTTAAAGAGAGAAATCTGTTTGAAGAGAAATTTAAGCAACACTG 1209
QY 191 AspCysIleGlyLysPheTyrlle---GlnGlyArgAlaPheThrLysGlu 207
```

[illegible]

```

Db      3083 GACAGACGACGGCATGAGCAGACGACGAGT 3113
      :::::|||||
RESULT 4
US-09-893-519A-146
; Sequence 146, Application US/09893519A
; Publication No. US20030027243A1
; GENERAL INFORMATION:
; APPLICANT: ANADYS PHARMACEUTICALS, INC.
; APPLICANT: THOMPSON, Craig
; APPLICANT: MOORE, Jeffrey
; APPLICANT: BURMAN, Ed T.
; APPLICANT: BRADLEY, John
; APPLICANT: DESILVA, Thamara
; APPLICANT: HARRIS, Sandra
; APPLICANT: KONARNITSKY, Svetlana
; APPLICANT: MENDILLO, Marc
; APPLICANT: MOORE, Daniel
; APPLICANT: MCCOY, Melissa
; APPLICANT: SANDERSON, Karen
; APPLICANT: HAO, Tariq
; APPLICANT: ZHU, Shuhao
; APPLICANT: LONG, Fan
; APPLICANT: DAVIDOV, Eugene
; TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
; FILE REFERENCE: 0342/1G548-US2
; CURRENT APPLICATION NUMBER: US/09/893,519A
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/215,164
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 60/224,457
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 146
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 146
; LENGTH: 2760
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: Human Genbank/AB015617.1
; DATABASE ENTRY DATE: 2000-01-06
; RELEVANT RESIDUES: (1)..(2760)
US-09-893-519A-146

Alignment Scores:
Pred. No.:      3,11e-05      Length:      2760
Score:          145.50      Matches:      95
Percent Similarity: 37.69%      Conservative: 81
Best Local Similarity: 20.34%      Mismatches:  164
Query Match:      4.69%      Indels:      127
DB:              11          Gaps:         20

US-09-890-475-1 (1-609) x US-09-893-519A-146 (1-2760)

QY      16 AlaAsnProLeuLeuGlnArgHisGlnSerGluGlnArgArgGluLeuProLysile 35
      :::::|||||
Db      1228 TCGAATGGTGTTCGTACTGTAGGAAGGAGAGAGAAATGAAGCAATG---GAAGTG 1284
      :::::|||||

QY      36 ValGluThrGluSerThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPhe 55
      :::::|||||
Db      1285 TATCGGAGCCATCTTCAATTTATGAAAAAATAAGATTGCCAGGTGAACACAG---GAGCTG 1341
      :::::|||||

QY      56 LeuLysSerIleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGln 75
      :::::|||||
Db      1342 TCCGAAGAAGCACAGAACTACTCGCCCTGCAGACAAAGCTAGAAACTCACAAPCCAG 1401
      :::::|||||

QY      76 PheAspAspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGlu 95
      :::::|||||
Db      1402 TTCTCAGATAGTAAACAGCACTTGAAGTGTGTGAAGGAGTCTTGTACTGCTAAG---GAG 1458
      :::::|||||

QY      96 SerAsnGlyValValLeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSerPro 115
      :::::|||||
Db      1459 CAGAGGGGTGCATCTCTG-----CAGACTGAGGTGCATGCT 1494
      :::::|||||

```

QY 529 aTyrGluHisLeuAlaProAsn-----SerTyrSerProGlyHisGly----- 543  
Db 2053 ATTACGCCACAGCTACCCGGGAGCTTCTGGTGGTTCACACCACTACTGGGGTCCAGG 2112  
QY 544 ----HisArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGlnArgHisProLe 562  
Db 2113 ACCTCACTGGGTTCATCCCATGGGTGGCCACCAACAGCTTGGGGC-----CATCCAA 2166  
QY 562 uGlnTyrSerPro-----ProIleHisGlyGlnGlnGlnLeuProTyrG1 577  
Db 2167 GCAAGGTGACCCACCCAGCCATGGGGTACCCCTTCAGGGCCATGCAGGGGTACCTCGAGG 2226  
QY 577 yIleGlnArgValTyrArgHisSerProSerGluGluArgTyrLeu-----GlyLe 594  
Db 2227 TAGCAGTATGGAGTCCGCAATAGCCCCAGGCTCTGAGGGGACAGCTTCTGGGGGACG 2286  
QY 594 uSerAsnGlnArgSerProArgSerAsnSer 604  
Db 2287 GACGGAGCAGGGCATGAGCAGAGACGACGAGT 2317

## RESULT 3

US-10-177-293-253  
; Sequence 253, Application US/10177293  
; Publication No. US20030124128A1  
; GENERAL INFORMATION:  
; APPLICANT: Lillie, James  
; APPLICANT: Glatt, Karen  
; APPLICANT: Zhao, Xumei  
; APPLICANT: Gannavarpu, Manjula  
; APPLICANT: Kamatkar, Shubhangi  
; APPLICANT: Mertens, Maureen  
; APPLICANT: Myer, Vic  
; APPLICANT: Wang, Youzheng  
; APPLICANT: Xu, Yongyao  
; APPLICANT: Hoersch, Sebastian  
; APPLICANT: Monahan, John  
; APPLICANT: Meyers, Rachel E.  
; APPLICANT: Bast Jr., Robert C.  
; APPLICANT: Hortobagyi, Gabriel N.  
; APPLICANT: Pusztai, Lajos  
; APPLICANT: Meric, Funda  
; APPLICANT: Sahin, Aysegul  
; APPLICANT: Mills, Gordon B.  
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,  
; TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER  
; FILE REFERENCE: MRI-038  
; CURRENT APPLICATION NUMBER: US/10/177,293  
; CURRENT FILING DATE: 2002-06-21  
; PRIOR APPLICATION NUMBER: US 60/299,887  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 60/301,572  
; PRIOR FILING DATE: 2001-06-27  
; PRIOR APPLICATION NUMBER: US 60/306,501  
; PRIOR FILING DATE: 2001-07-18  
; PRIOR APPLICATION NUMBER: US 60/325,002  
; PRIOR FILING DATE: 2001-09-25  
; PRIOR APPLICATION NUMBER: US 60/362,585  
; PRIOR FILING DATE: 2002-03-05  
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx  
; PRIOR FILING DATE: 2002-05-14  
; NUMBER OF SEQ ID NOS: 506  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 253  
; LENGTH: 4620  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-177-293-253

Alignment Scores:  
Pred. No.: 3.26e-05 Length: 4620  
Score: 148.50 Matches: 150  
Percent Similarity: 34.88% Conservative: 98  
Best Local Similarity: 21.10% Mismatches: 227

Query Match: 4.78% Indels: 237  
DB: 14 Gaps: 35  
US-09-890-475-1 (1-609) x US-10-177-293-253 (1-4620)  
QY 3 AsnTyr-----ProProThrValAlaAlaGlnProThrThrThrAlaAsn 17  
Db 1365 AATTACAGAGAGAGGAGATCTTAGAACAGAGCATCAGATCCAACTCCA----- 1418  
QY 18 ProLeuLeuGlnArgHisGlnSerGluGlnArgArgGluLeuProLysIleValGlu 37  
Db 1419 CCCCAAGTATCTCGTCACAAATCACATATCGTAATCGAGAACACATTTGTCTACTACGG 1478  
QY 38 ThrGluSerThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLys 57  
Db 1479 ACAGCATCACTG-----GTTACGAGGCAAAATGCAAGAACATGAG 1517  
QY 58 SerIleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAsp 77  
Db 1518 CAGGACTCTGAGCTTAGAGAACAAATGCTGGCTATTAAGCGAATGAGCGCACACAT-- 1574  
QY 78 AspLeuGlnLysHisIleGluSerIleGluSerIleAspSerLysLeuGluSerAsn 97  
Db 1575 -----CAAAAGCACTGATGACTCTGGAACCAAGCTAAAGGTGAGATGAGTAACAT 1628  
QY 98 GlyValValLeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSerProArg 117  
Db 1629 CCGCTCAGATTA-----GACAAAGATCTTGAACCTCAGCGT 1664  
QY 118 AsnAsnValSerValGluThrThrValThrValSerGlnProSerGlnGluIleValPro 137  
Db 1665 AACAAATTTTCTGTCAGAAATGGAGAACTTATCAAGAAACACCAGCGCTGTATGGAGAAA 1724  
QY 138 GluThr-----SerAsnLysProGluGlyArgMetCysGluLeuMetCysSer 154  
Db 1725 GAGGCTAAGTATGTCATGAA-----GAG 1751  
QY 155 LysGlyLeuArgLysTyrIleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGluGlu 174  
Db 1752 AAAAAATTTTCAGCAACATATT-----CAGGCCCAACACAGAGAAAGAA 1793  
QY 175 IleProSerAlaLeu-----LysLeuAlaLysGluProAlaLys 187  
Db 1794 CTGAATAGTTTCTCGAGTCCAGAAAGAGAGATATAACTTCGAAAGAGCAGCTTAAA 1853  
QY 188 PheValLeuAspCysIleGlyLysPheTyrLeuGlnGlyArg-----AlaPhe 204  
Db 1854 GAGGAGCTAAAT--GAAACACCAGAG-TACCCCAAAAAAGAAAAACAGGAGTGGCTTTC 1909  
QY 205 ThrLysGluSer-----ProMetSerSer 212  
Db 1910 AAAGCAGAGGAGATATACAGCATTTCCAGCAGAGAGAGTAACCTTCTTCGACG 1969  
QY 213 AlaArgGlnValSerLeuLeuIleLeuGluSerPheLeuLeuMetProAspArgGlyLys 232  
Db 1970 TCAAGACAAATA-----CCTAGAGCTGGAATG 1996  
QY 233 GlyLysValLysIleGluSerTrpIleLysaspGluAlaGluThrAlaAlaValAlaTrp 252  
Db 1997 CCGTCGCTTCAAGAGAGAAAT-----GTTACTTGG 2026  
QY 253 ArgLysArgLeuMetThr-----GluGlyGlyLeuAlaAlaAlaGluLysMetAsp 269  
Db 2027 GCGTCATACTTAGACGAGGACCTTGTCCAGGGAGGAGTT-----AAACAAAGACAGAC 2080  
QY 270 AlaArgGlyLeuLeuLeuValAlaCysPheGlyValProSer----- 284  
Db 2081 TCAGAAAGGACTTAGA-----GCATGCCATGTACTCCGACGAGCATGAATCTATGCA 2131  
QY 285 -----AsnPheArgSerThrAspLeuLeu-AspLeuIleArgMe 297  
Db 2132 AGAAGTGGAGTTCGGCCACCTCAACACAAATTCAGAGAGTGGCTGTGAGTTCAGATT 2191

CURRENT APPLICATION NUMBER: US/09/291,417A  
 CURRENT FILING DATE: 1999-04-13  
 EARLIER APPLICATION NUMBER: US 60/081,784  
 EARLIER FILING DATE: 1998-04-14  
 NUMBER OF SEQ ID NOS: 147  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 20  
 LENGTH: 3824  
 TYPE: DNA

ORGANISM: Mammalian (Human) SULU3

S-09-291-417-20

Alignment Scores:  
 Seq. No.: 2,43e-05 Length: 3824  
 Score: 148.50 Matches: 150  
 Percent Similarity: 34.88% Conservative: 98  
 Best Local Similarity: 21.10% Mismatches: 227  
 Query Match: 4.78% Indels: 237  
 Gaps: 35

US-09-890-475-1 (1-609) x US-09-291-417-20 (1-3824)

```

Y 3 AsnTyr-----PropThrValAlaAlaGlnProThrThrThrAlaAsn 17
b 569 AATTACAGAGAGAGGAGAGTCTAGAACAGAGCATCAGATCCACAATCTCCA----- 622
Y 18 ProLeuLeuGlnArgHisGlnSerGluGlnArgArgGluLeuProLysIleValGlu 37
b 623 CCCAAGATCTCGTCACAAATCACACTATCGTAATCGAGAACACTTCTCTACTATACG 682
Y 38 ThrGluSerThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLys 57
b 683 ACAGCATCACTG-----GTTACGAGGCCAATGCAAGAACATGAG 721
Y 58 SerIleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAsp 77
b 722 CAGGACTCTGAGCTTAGAGAACAAATGCTCGGTATAGCGAATGAGGCGACAACAT--- 778
Y 78 AspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsn 97
b 779 -----CAAAGCAACTGATGACTCTCGAATAAACAAGCTAAGGCTGAGATGAACAT 832
Y 98 GlyValValLeuAlaAlaArgAsnAsnPheHisGlnProMetLeuSerProProArg 117
b 833 CGCCTCAGATTA-----GACAAAGATCTTGAACACTCAGCGT 868
Y 118 AsnAsnValSerValGluThrThrValThrValSerGlnProSerGlnGluIleValPro 137
b 869 AACCAATTTGCTGCAGAAATGGAGAAACTTATCAAGAAACACCAGGCTGCCAATGGAGAA 928
Y 138 GluThr-----SerAsnLysProGluGlyGlyArgMetCysGluLeuMetCysSer 154
b 929 CAGGCTAAGTGATGTCCAATGAA-----GAG 955
Y 155 LysGlyLeuArgLysTyrIleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGluGlu 174
b 956 AAAAAATTCAGCAACATATT-----CAGGCCCAACAGAGAAGAA 997
Y 175 IleProSerAlaLeu-----LysLeuAlaLysGluProAlaLys 187
b 998 CTGAATAGTTTCTCGAGTCCCAAGAGAGAGTATAACTTCGAAAGAGCAGCTTAAA 1057
Y 188 PheValLeuAspCysIleGlyLysPheTyrLeuGlnGlyArg-----AlaPhe 204
b 1058 GAGGAGCTAAAT-----GAAACACAGAG-TACCCCAAAAAAGAAAAACAGAGTGGCTTTC 1113
Y 205 ThrLysGluSer-----ProMetSerSer 212
b 1114 AAAGCAAGAGAGATATACAGCATTTCCAGACAGAGAGAGACTTAACCTTCTTCGAGC 1173
Y 213 AlaArgGlnValSerLeuLeuIleGluSerPheLeuLeuMetProAspArgGlyLys 232
b 1174 TCMAAGACAATA-----CCTAGAGTGGAAATG 1200

```

```

QY 233 GlyLysValLysIleGluSerTrpIleLysAspGluAlaGluThrAlaAlaValAlaTyr 252
Db 1201 CGTCGCTTCAAGAGAAGAA-----GTTACTTGG 1230
QY 253 ArgLysArgLeuMetThr-----GluGlyGlyLeuAlaAlaAlaGluLysMetAsp 269
Db 1231 GCGTCATAACTTAGAGCAGGACCTTGTCAAGGAGGAGTT-----AAACAAAGACAGAC 1284
QY 270 AlaArgGlyLeuLeuLeuValAlaCysPheGlyValProSer----- 284
Db 1285 TCAGAAGGAGCTTAGA-----GCATGCCATGCTACTCCGACAGCATGAATCTATGCA 1335
QY 285 -----AsnPheArgSerThrAspLeuLeuAspLeuIleArgMe 297
Db 1336 AGAAGTGGAGTTCGCCACCTCAACACATTCAGAGATGCGCTGTGAGTTGATCAGATT 1395
QY 297 tSerGlySerAsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetVa 317
Db 1396 ACAGCATCAAACTGAGCTCACTAACGAGCTGGAATATATAAGCGAAGA----- 1444
QY 317 lSerGlyIleValGluSerSerIleLysArgGlyMetHisIleGluAlaLeuGluMetVa 337
Db 1445 -----GACGAGAACTAAGACGA---AAGCATGTC----- 1471
QY 337 lTyrThrPheGlyMetGluAspLysPheSerAlaAlaLeuValLeuThrSerPheLeuLys 357
Db 1472 -----ATGGAAGTTCGACACACAGCCTAAGAGTTTGAAGTCTAAAGAACTCCA 1518
QY 357 sMetSerLysGluSerPheGluArgAlaLysArgLysAlaGlnSerProLeuAla----- 375
Db 1519 AATAAAAGACGAGTTTCAGGATACCTGCAAAATCCAAACACAGACAGTACAAAGCATTAAG 1578
QY 376 -----PheLys 377
Db 1579 AATCACCCTGCTGGAGACTACACCAAGAGTGCAGACAAAGCTGTTCTGAAACGCTCAA 1638
QY 377 sGluAlaAlaThrLysGlnLeuAlaValLeuSer-----SerValMetGlu 392
Db 1639 GGAGGAACAGACCCGAAATTAGCTATCTTGCTGAGCAGCATGATCATCAGCATTAATGA 1698
QY 392 nCysMetGluThrHis-----LysLeuAspProAlaLysGluLeuProGlyTyrGlnIle 410
Db 1699 AATGCTCTCCACACAAAGCCCTCGTGTGGATGAAGCAGACGAGGAGTGCCAGGTTT 1758
QY 410 eLysGluGlnIle-----ValSerIle 417
Db 1759 GAAGATGACGCTGCAGCAGGAACTGGAGCTGTTGAATGCGTATCAGAGCAAAATCAAGAT 1818
QY 417 uGluLysAspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSerLeuSerIle 437
Db 1819 GCAAGCTGAGGCA---CAACATGATCGAGAGCTTCGAGAGCTTGACAGAGGCTCCCT 1875
QY 437 u-----MetGluGluAlaAlaLeuAlaLysArgMetTyrAs 449
Db 1876 CCGGAGGCGCACTCTTAGACAAAGATTCAGAGAGAGAGTGTGGCT-----TTGCAGAA 1929
QY 449 nGlnGlnIleLysArgProArgLeuSerProMetGluMetProProValThrSerSerSe 469
Db 1930 TGAGCGCACAGAACGAATACGA----- 1951
QY 469 rTyrSerProIleTyrArgAspArgSerPheProSerGlnArgAspAspGlnAspGln 489
Db 1952 -----AGCCTGTTGGAACGCTCAAGCCAGACA 1977
QY 489 uIleSerAlaLeuValSerSerTyrLeuGlyProSerThrSerPheProHisArgSerAr 509
Db 1978 GATTGAAGCTTTTGAAGCTGAAAGCATGAGACTAGGTTTTAGTATATATGCTCTTCTTAA 2037
QY 509 gArgSerProGluTyrMetValProLeuProHisGlyGlyLeuGlyArgSerValTyrAl 529
Db 2038 TCTCTCCCTCGAG-----GC 2052

```

[illegible]

GenCore version 5.1.6  
Copyright (C) 1993 - 2003 CompuGen Ltd.

M protein - nucleic search, using frame\_plus\_p2n model

un on: August 25, 2003, 12:30:10 ; Search time 384 Seconds

(without alignments)

3565.447 Million call updates/sec

file: US-09-890-475-1

effect score: 3104

sequence: 1 MGNYPPTVAAQPTTANPL.....RYLGLSNQSPRSNSLDPK 609

coring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Fgapop 10.0, Fgapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

searched: 1517243 seqs, 1124081882 residues

total number of hits satisfying chosen parameters: 3034486

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

MODEL=frame+p2n.model -DRV=xlh

Q=/cgn2\_1/USPTO.spool/US09890475/runat\_14082003\_090354\_19703/app\_query.fasta\_1.775

DB=Published Applications NA -OFMT=fastcap -SUFFIX=rnpb -MINMATCH=0.1

LOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62

TRANS=human40.cdd -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100

THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0

MAXLEN=200000000 -USER=US09890475@cgn\_1\_1238@runat\_14082003\_090354\_19703

NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100

LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5

FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq\*

2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq\*

3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq\*

4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq\*

5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq\*

6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq\*

7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq\*

8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq\*

9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq\*

10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq\*

11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq\*

12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq\*

13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq\*

14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq\*

15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq\*

16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq\*

17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query				Description
No.	Score	Match	Length	ID	
1	234.5	7.6	1524	10	US-09-938-842A-917 Sequence 917, App

2	148.5	4.8	3824	11	US-09-291-417-20 Sequence 20, Appl
3	148.5	4.8	4620	14	US-10-177-293-253 Sequence 253, App
4	145.5	4.7	2760	11	US-09-893-519A-146 Sequence 146, App
5	144	4.6	2756	10	US-09-960-253-159 Sequence 159, App
6	140	4.5	3718	10	US-09-917-800A-1471 Sequence 1471, App
7	134	4.3	2526	10	US-09-938-842A-1242 Sequence 1242, Ap
8	132.5	4.3	5142	14	US-10-097-340-42 Sequence 42, Appl
9	131	4.2	9775	9	US-09-751-962-1 Sequence 1, Appli
10	130.5	4.2	7393	14	US-10-037-270-372 Sequence 372, App
11	129	4.2	3324	14	US-10-037-270-1020 Sequence 1020, Ap
12	129	4.2	9549	14	US-10-204-887-31 Sequence 31, Appl
13	129	4.2	3309400	10	US-09-738-626-1 Sequence 1, Appli
14	127.5	4.1	10432	11	US-09-919-172-97 Sequence 97, Appl
15	127.5	4.1	10432	11	US-09-919-039-20 Sequence 20, Appl
16	127.5	4.1	10432	14	US-10-084-817-349 Sequence 349, App
17	127.5	4.1	12515	11	US-09-918-624B-48 Sequence 48, Appl
18	127.5	4.1	12632	13	US-10-071-766-49 Sequence 49, Appl
19	126.5	4.1	9934	9	US-09-751-962-2 Sequence 2, Appli
20	124.5	4.0	3284	14	US-10-128-714-7305 Sequence 7305, Ap
21	124.5	4.0	3379	14	US-10-128-714-6305 Sequence 6305, Ap
22	124.5	4.0	5379	14	US-10-128-714-5305 Sequence 5305, Ap
23	124	4.0	1278	14	US-10-156-761-5245 Sequence 5245, Ap
24	124	4.0	9025608	14	US-10-156-761-1 Sequence 1, Appli
25	122.5	3.9	6604	10	US-09-880-107-1748 Sequence 1748, Ap
26	122.5	3.9	7100	10	US-09-932-183A-1 Sequence 1, Appli
27	121	3.9	3811	14	US-10-177-293-105 Sequence 105, App
28	121	3.9	14800	10	US-09-954-456-1601 Sequence 1601, Ap
29	120	3.9	11220	10	US-09-861-289-32 Sequence 32, Appl
30	120	3.9	11220	10	US-09-860-846-32 Sequence 32, Appl
31	120	3.9	11220	11	US-09-988-384B-32 Sequence 32, Appl
32	120	3.9	11220	11	US-09-836-821-32 Sequence 32, Appl
33	120	3.9	36778	10	US-09-861-289-5 Sequence 5, Appli
34	120	3.9	36778	10	US-09-860-846-5 Sequence 5, Appli
35	120	3.9	36778	11	US-09-836-821-5 Sequence 5, Appli
36	120	3.9	37948	11	US-09-988-384B-5 Sequence 3717, Ap
37	119	3.8	3075	10	US-09-880-107-3717 Sequence 102, App
38	119	3.8	3080	14	US-10-171-581-102 Sequence 1921, App
39	119	3.8	7869	10	US-09-954-456-1921 Sequence 297, App
40	119	3.8	7923	10	US-09-864-864-297 Sequence 11, Appl
41	119	3.8	7923	14	US-10-239-058-11 Sequence 1602, Ap
42	118.5	3.8	3388	10	US-09-954-456-1602 Sequence 245, App
43	118.5	3.8	3388	10	US-09-967-768A-245 Sequence 988, App
44	118.5	3.8	3388	10	US-09-954-531-988 Sequence 1382, Ap
45	118.5	3.8	3388	10	US-09-954-531-1382

ALIGNMENTS

RESULT 1

US-09-938-842A-917

; Sequence 917, Application US/09938842A

; Patent No. US20020160378A1

; GENERAL INFORMATION:

; APPLICANT: Harper, Jeff

; APPLICANT: Kieps, Joel

; APPLICANT: Wang, Xun

; APPLICANT: Zhu, Tong

; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING

; FILE REFERENCE: SAME, AND METHODS OF USE

; CURRENT FILING DATE: 2001-08-24

; PRIOR APPLICATION NUMBER: US 60/227,866

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: US 60/264,647

; PRIOR FILING DATE: 2001-01-16

; PRIOR APPLICATION NUMBER: US 60/300,111

; PRIOR FILING DATE: 2001-06-22

; NUMBER OF SEQ ID NOS: 5379

; SEQ ID NO 917

; LENGTH: 1524

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

```

3790 AAGCTGGAGGAGCTCGGCTCTGCGAGCAGAGACAGCCAGCAACAGTGCAGAGCTG 3849
140 SerAsnLysPro---GluGlyGlyArgMetCys----- 149
3850 CAGAACCCAGCTGCTCTCGGAGGAGGTGAGAGCCCTCCGGAGGAGGCTGAGAAA 3909
150 GluLeuMetCysSerLysGlyLeuArgLysTyrIleTyrAlaAsnIleSerAspGlnAla 169
3910 CAGCGGTGGCTTCAGAGAACCTTGGCCAG-----GAGTGCACCTCACAGGCT 3957
170 LysLeuMetGluGluIleProSerAlaLeuLysGlu----- 184
3958 GAGCGTGGAGGAGCTGGCCAGAAATTGAAGCGGTGGCAGGAGAGTTCTTCCAGAAA 4017
184 ----- 184
4018 GAGCAGGCCCTCTCCACCTCGAGCTCGAGCAGACACAGCAGCCCTGGTGAAGTAG 4077
185 -----ProAlaLysPheValLeuAspCysIleGlyLysPheTyrLeuGlnGlyArgArg 202
4078 CTGCTGCCAGCTAAGCACTCTGCCAG-----CAGTGCAGGCCGAGCAG 4122
203 AlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuIle----- 220
4123 GCCGCTGCCGAGAAACCCACCGTGGAGAGCTGGAGCAGAGCAAGCAGCCGCTGGGGGA 4182
221 LeuGluSerPheLeuMetProAspArgGlyLysLys---ValLysIleGluSer 239
4183 CTGCGGGCAGAGCTGCTCGGGGCCAGCGGAGCTTGGGAGCTGAATCTCTCGGGCAG 4242
240 TrpIleLysAspGlu-----AlaGluThrAlaAlaValAla 251
4243 AAGGTGGCAGAGCAGGAGCGAAACAGCTCAGCAGCTGGCGGCAGAGAGCCAGCTATGCA 4302
252 TrpArgLysArgLeuMetThrGluGly---GlyLeuAlaAlaGluLysMetAspAla 270
4303 GAGCAGCTGAGCATGCTGAAGAGAGCGGCATGGCTGTGCGCAGAGGAGAAC----- 4353
271 ArgGlyLeuLeuLeuValAlaCysPheGlyValProSerAsnPheArgSerThrAsp 290
4354 CGGGGGCTG-----GGTGAGCGGCCCAACCTTGGCGCGCAGTTT 4392
291 Leu-----LeuAspLeuIleArgMetSerGlySerAsnIleLeuAlaGlyAlaLeu 307
4393 CTGGAAGTGGAGTTGGACCAAGCCGCGGAAAGATATGTCACAGAGTTGGCAGCGGTACGT 4452
308 LysArgSerGlnPheLeuValProMetValSerGlyIleValGluSerIleLys--- 326
4453 GCTGATGCTGAGACCGCTCTGGCTGAGTGCAGCGAGAGCAGCAGAGCACTGCCCGGAG 4512
327 -----ArgGlyMetHisIleGluAlaLeuMetValTyr 338
4513 CTGAGAGTGATGACTGCCAAGTATGAGGTGCCAAGGTCAAGGTCTCTGAGGAGAGGCAG 4572
339 ThrPheGlyMetGlu----- 343
4573 CGGTTCCAGGAAGAGAGCAGAAACTCACTGCCAGGTGGAAGAACTGAGTAGAAGAACTG 4632
344 ---AspLysPheSerAlaAlaLeuValLeuThrSerPheLeuLysMetSerLysGluSer 362
4633 GCTGACTCTGACCAAGCCAGCAAGGTGCAGCAGCAGAAAGCTGAAGGTGTCCAGGCTCAG 4692
363 PheGluArgAlaLysArgLysAlaGlnSerProLeuAlaPheLysGluAlaAlaThrLys 382
4693 GGAGCGGAGCCAGCAGAGGAGGAGGAG-----CGCTTCAGGCCCGCCAGCTGAATGAA 4743
383 GlnLeuAlaValLeuSerSerValMetGlnCysMetGluThrHisLysLeuAsp----- 400
4744 CTGCAAGCCAGTTGAGCCAGAGGAGCAGCAGCTGAGCAGCTATAGCTGCAGATGGAG 4803
401 -----ProAlaLysGluLeuProGlyTrpGlnIleLysGluGlnIle 414
4804 AAAGCCAAACACATTATGATGCCAAGAGCAGCAGAAACCAAGAGCTGCAGGAGCAGCTG 4863

```

```

415 ValSerLeuGluLysAspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSer 434
4864 CGGAGGCTGGAG-----CAGCTGCAGAAAGAAACAAAGAG----- 4899
435 LeuSerLeuMetGluGluAlaAlaLeuAla---LysArgMetTyrAsnGlnGlnIleLysArg 454
4900 -----CTGCGAGCTGAGCTGAACGGCTGGGCCATGAGCTACAGCAGGCTGGGCTGAAG 4953
454 gProArgLeuSerProMetGluMetProProValThrSerSerSerTyrSerProIleTyr 474
4954 ACCAAGG-----AGGCTGAACAGAGCTGCCGCCACCTTACTGCCCCAGGTGGCAGCCTG 5007
474 ArgAspArgSerPheProSerGlnArgAspAspGlnAspGluIleSerAla----- 492
5008 GAGGACAGS---TTGCCCATGCAGACAGCAGCTTCGAGAGCTGGGCAATTCAGGCTG 5064
493 -----LeuValSerSerTyrLeuGlyProSer----- 501
5065 GCAACTGATGCTTTAAAGAGCGGTGAGCCCGCAGGCTAAGCCCGCAGCTGGACTTGAAT 5124
502 -ThrSerPheProHisArgSerArgSerArgSerProGluTyrMetValPro----- 517
5125 GACAGCTGGATCTGAGCTGCGAGGAGGAGCCCACTCAGTATCACCAGCAGCTGCT 5184
518 -LeuProHisGlyGlyLeuGlyArgSerValTyrAlaTyrGluHisLeuAlaProAsnSe 537
5185 CGTACCCAGCAGACCGCAGCCAGCGCTCCCTGGAGAACCAAGCTCCTATCTCCAGCGC 5244
537 rTyrSerPro-GlyHisGlyHisArgLeuHisArgGlnTyrSerProSerLeuValHisG 557
5245 CTGCCCCCAGGT-----AGATCCCTGGAGAGTCTCTACTTCACT 5286
557 LyGlnArgHisProLeuGlnTyrSerProIleHisGlyGlnGlnLeuProTyrG 577
5287 CC-----CATCCCTGCTCGAGTCAAGGCCCTGGAGAGCAGCTGGACTCCCTGGG- 5339
577 LysLeuArgVal----- 581
5340 -----AGACGCTCTTCTGGACTCGGCTCGTAAGACCGCTCGCTCGGCGCACACAG 5394
582 -----TyrArgHisSerProSerGluGluArgTyrLeuGlyLeuSerAsnGlnArg 598
5395 CAGTATCATCAATCACCATGACCAAGAGAGTAGATGTGGAAGAGCCAGACAGCGC 5450

```

Search completed: August 25, 2003, 13:28:25  
Job time : 274 secs



## ATTORNEY/AGENT INFORMATION:

NAME: PITCHER ESQ, EDMUND R  
 REGISTRATION NUMBER: 27,829  
 REFERENCE/DOCKET NUMBER: MTP-013  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 617/248-7000  
 TELEFAX: 617/248-7100  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 6306 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 FEATURE:

NAME/KEY: CDS  
 LOCATION: 1..6306  
 PUBLICATION INFORMATION:  
 AUTHORS: COMPTON, DUANE A  
 AUTHORS: SZILAK, ILLYA  
 AUTHORS: CLEVELAND, DON W  
 TITLE: PRIMARY STRUCTURE OF NIMA, AN INTRANUCLEAR  
 TITLE: PROTEIN THAT DEFINES A NOVEL PATHWAY FOR  
 TITLE: SEGREGATION OF PROTEINS AT MITOSIS  
 JOURNAL: JOURNAL OF CELL BIOLOGY  
 VOLUME: 116  
 PAGES: 1395-1408  
 DATE: MAR-1992  
 JS-08-195-487-3

## Alignment Scores:

pred. No.: 0.059 Length: 6306  
 score: 127.00 Matches: 161  
 Percent Similarity: 34.61% Conservative: 102  
 Best Local Similarity: 21.18% Mismatches: 275  
 Query Match: 4.03% Indels: 222  
 Gaps: 35

JS-09-890-475-1 (1-609) x US-08-195-487-3 (1-6306)

2y 9 AlaAlaGlnProThrThrThrAlaAsnProLeuLeuGlnAArgHisGlnSerGlu----- 26  
 Db 3331 GCTGCTGCAGGACAGAGCCACAGGCCCAAGCTGGAAGCACTGGCGGACAGGTGAGC 3390  
 2y 27 -----GlnAArgArgGlu-----LeuProLysIleVal 36  
 Db 3391 AAGCTGGAACAGCAATGCCAGAGCAGCAGGACGAGCTGACAGCCTGGAACGACGCTC 3450  
 2y 37 GluThrGluSerThrSer-----MetAspIleThrIleGlyGln 49  
 Db 3451 GAGGCTGAGCGGGCTCCCGGGCTGAGCGGGACAGTGTCTGTGAGACTCTGCGAGGCCAG 3510  
 2y 50 -----SerLysGlnProGlnPheLeuLysSerIleAsp 60  
 Db 3511 TTAGAGGAGAGAGGCCAGGAGCTAGGCGACAGTCAGAGTGCCTTAGCCTCGGCCAACGG 3570  
 2y 61 GluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAspLeuGln 80  
 Db 3571 GAGTTGGTGTCTTCGGACCAAGGTTACAGACCACAGCAGGCTGAGATGAGTGGAG 3630  
 2y 81 LysHisIle-----GluSerIleGluAsnAlaIleAspSerLysLeu 94  
 Db 3631 GCCCAGGTGGCCCGGGCCGCAAGAGGCTGAGAGGAAATAAGCCTCATCAGCAGCTTG 3690  
 2y 95 GluSerAsnGlyValValLeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSer 114  
 Db 3691 GAGGAGGAGGTTCATCCTG-----AATGCCAGGTTCCTGGAG 3729  
 2y 115 ProProArgAsnAsnValSerValGluThrThrValThrVal-----SerGlnProSerGln 133  
 Db 3730 AAGGAGGGGAGAGCAAGGAGTTGAAGCGGCTGTGATGTCGCGAGTCAGAGAAGACCCAG 3789  
 2y 134 GluIle-----ValProGluThr 139

Db 3790 AAGCTGGAGGAGAGCTGCGCTGCTGTCAGGAGAGAGACAGCCAGCAACAGTCGCCAGAGCTG 3849  
 Qy 140 SerAsnLysPro-----GluGlyGlyArgMetCys----- 149  
 Db 3850 CAGAACCGCAGCTCTGCTCTGCGGGAGGAGGTGCCAGAGCTCCGGAGGGAGGCTGAGAAA 3909  
 Qy 150 GluLeuMetCysSerLysGlyLeuArgLysTyrIleTyrAlaAsnIleSerAspGlnAla 169  
 Db 3910 CAGCGGTGGCTTCAGAGAACCTGCGGCAG-----GAGCTGACCTCACAGCT 3957  
 Qy 170 LysLeuMetGluGluIleProSerAlaLeuLysLeuAlaLysGlu----- 184  
 Db 3958 GAGCGTGGAGGAGCTGGGCCAAGATTGAAGCGTGGCAGAGAGAGTTCTTCAGAAA 4017  
 Qy 184 ----- 184  
 Db 4018 GAGCAGGCCTCTCCACCTGTCAGCTCGAGCACACACAGCCACACAGGCGCTGTCGAGTGAG 4077  
 Qy 185 -----ProAlaLysPheValLeuAspCysIleGlyLysPheTyrLeuGlnGlyArgArg 202  
 Db 4078 CTGCTGCCAGCTAAGCACCTCTGCCAG-----CAGCTGACGCCCGCAGCAG 4122  
 Qy 203 AlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuIle----- 220  
 Db 4123 GCCGCTGCCAGAGAACGCCACCGTCGAGAGCTGAGAGCAGACAGCAGCGCGCTGGGGA 4182  
 Qy 221 LeuGluSerPheLeuLeuMetProAspArgGlyLysGlyLys-----ValLysIleGluSer 239  
 Db 4183 CTGCGGCGAGAGCTGCTGCGGGCCAGCGGAGCTGGGGAGCTGATTCCTCTCGCGCAG 4242  
 Qy 240 TrpIleLysAspGlu-----AlaGluThrAlaAlaValAla 251  
 Db 4243 AAGGTGGCAGAGCAGGACCGAACAGCTCAGCAGCTGCGGGCAGAGAGCCAGCTATGCA 4302  
 Qy 252 TrpArgLysArgLeuMetThrGluGly-----GlyLeuAlaAlaGluLysMetAspAla 270  
 Db 4303 GAGCAGCTGAGCATGCTGAAGAAGCGCATGCGCTGCTGTCGAGAGGAGAAC----- 4353  
 Qy 271 ArgGlyLeuLeuLeuValAlaCysPheGlyValProSerAsnPheArgSerThrAsp 290  
 Db 4354 CGGGGGCTG-----GGTGACGGCGCAACCTTGGCGCGGAGTTT 4392  
 Qy 291 Leu-----LeuAspLeuIleArgMetSerGlySerAsnGluIleAlaGlyAlaLeu 307  
 Db 4393 CTGGAAGTGGAGTTGACCGAGCCCGGAAAGATATGCCAAGAGTTGGCAGCCCTACGT 4452  
 Qy 308 LysArgSerGlnPheLeuValProMetValSerGlyIleValGluSerSerIleLys--- 326  
 Db 4453 GCTGATGCTGAGACCGCTCTGGGTGAGTGCAGCGAGAGACACAGAGCACTGCCCGGAG 4512  
 Qy 327 -----ArgGlyMetHisIleGluAlaLeuGluMetValTyr 338  
 Db 4513 CTGAGGTGATGATCTGCCAAGTATGAGGTGCCAGGTCCAGGTCTTGAGGAGGAGCGAG 4572  
 Qy 339 ThrPheGlyMetGlu----- 343  
 Db 4573 CGGTTCCAGGAGAGAGGCGAGAACTCACTGCCAGGTGGAAGAACTGAGTAAGAACTG 4632  
 Qy 344 ---AspLysPheSerAlaAlaLeuValLeuThrSerPheLeuLysMetSerLysGluSer 362  
 Db 4633 GCTGACTCTCCACCAAGCCAGAGGTGCGAGCAGCAGAGCTGCTGAGGCTGCTGAGGCTCAG 4692  
 Qy 363 PheGluArgAlaLysArgLysAlaGlnSerProLeuAlaPheLysGluAlaAlaThrLys 382  
 Db 4693 GGAGCGGAGAGCCAGCAGGAGGCGCCAG-----CGCTTCCAGGCCCGCCTGAATGAA 4743  
 Qy 383 GlnLeuAlaValLeuSerSerValMetGlnCysMetGlnThrHisLysLeuAsp----- 400  
 Db 4744 CTGCAAGCCCGAGTTGAGCCAGAGAGGAGCGAGCTGAGCACTATAAGCTGAGATGGAG 4803  
 Qy 401 -----ProAlaLysGluLeuProGlyTrpGlnIleLysGluGlnIle 414

140 SerAsnLysPro---GluGlyGlyArgMetCys----- 149  
 3850 CAGAACGACAGCTCTGCTCTCGCGGAGGAGGTGCAGAGCCCTCGCGGAGGAGGCTGAGAAA 3909  
 150 GluLeuMetCysSerLysGlyLeuArgLysTyrIleTyrAlaAsnIleSerAspGlnAla 169  
 3910 CAGCGGTGGCTTCAGAGAACCTTGGCGCAG-----GAGTGCACCTCAGAGCT 3957  
 170 LysLeuMetGluGlnIleProSerAlaLeuLysLeuAlaLysGlu----- 184  
 3958 GAGCGTCCGAGGAGCTGGGCCAAGATTGAAGCGTGGCAGAGAGATTCTTCCAGAAA 4017  
 184 ----- 184  
 4018 GAGCAGGCCCTCTCCACCTCGAGCTCGAGCACACAGCAGCAGCCCTGTGTAGTGAG 4077  
 185 -----ProAlaLysPheValLeuAspCysIleGlyPheTyrLeuGlnGlyArg 202  
 4078 CTGCTGCAGCTAAGCACCTTCTGCCAG-----CAGCTGCAGCCGAGCAG 4122  
 203 AlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuIle----- 220  
 4123 GCCCTCCGAGAGAACGCCACCGTAGGAGCTGGAGCAGCAGCAGCCCGCTGGGGA 4182  
 221 LeuGluSerPheLeuLeuMetProAspArgGlyLysLys---ValLysIleGluSer 239  
 4183 CTGCGGCGAGAGCTGCTCGCGGCCAGCGGAGCTTGGGAGCTGATCTCTCGGGCAG 4242  
 240 TrpIleLysAspGlu-----AlaGluThrAlaAlaValAla 251  
 4243 AAGGTGGCAGCAGCAGAGCAGCAGCTCAGCAGCTCGCGGCGCAGAGCCAGCTATGCA 4302  
 252 TrpArgLysArgLeuMetThrGluGly---GlyLeuAlaAlaAlaGlyLysMetAspAla 270  
 4303 CAGCAGCTGACATGCTGCTGAGAGCGCATGGCTGCTGCGCAGCAGAGAAC----- 4353  
 271 ArgGlyLeuLeuLeuValAlaCysPheGlyValProSerAsnPheArgSerThrAsp 290  
 4354 CGGGGGCTG-----GGTGAGCGGGCCACCTTGGCGCGCAGTTT 4392  
 291 Leu-----LeuAspLeuIleArgMetSerGlySerAsnGluIleAlaGlyAlaLeu 307  
 4393 CTGGAAGTGGAGTTGGACCGCGCGGAGAAAGATATGTCACAGAGTTGGCAGCGTACGT 4452  
 308 LysArgSerGlnPheLeuValProMetValSerGlyIleValGluSerSerIleLys--- 326  
 4453 GCTGATGCTGAGACCGCTGCTGAGTGCAGCGCAGCAGCAGCAGCACTGCGCGGAG 4512  
 327 -----ArgGlyMetHisIleGluAlaLeuGluMetValTyr 338  
 4513 CTGGAGGTGATGATCGCCAAAGTATGAGGTGCCAAGTCAAGGTCCTGGAGGAGCGCAG 4572  
 339 ThrPheGlyMetGlu----- 343  
 4573 CGGTTCCAGGAAGAGGAGCAGAACTCACTGCCAGGTGGAAGAACTGAGTAAGAAACTG 4632  
 344 ---AspLysPheSerAlaAlaLeuValLeuThrSerPheLeuLysMetSerLysGluSer 362  
 4633 GCTGACTCTGACCAAGCAGCAGAGGTGCGCAGCAGAGCTGAAGGCTGTCCAGGCTCAG 4692  
 363 PheGluArgAlaLysArgLysAlaGlnSerProLeuAlaPheLysGluAlaAlaThrLys 382  
 4693 GGAGCGGAGAGCAGCAGAGGCCAG-----CGCTCCAGGCCAGCTGAATGAA 4743  
 383 GlnLeuAlaValLeuSerSerValMetGlnCysMetGluThrHisLysLeuAsp----- 400  
 4744 CTGCAAGCCCACTTGAAGCCAGCAGCAGCAGCAGCTGAGCACTATAAGCTGCAGATGGAG 4803  
 401 -----ProAlaLysGluLeuProGlyTyrGlnIleLysGluGlnIle 414  
 4804 AAAGCCAAACATTTATGATCCCAAGAGCAGCAGCAGCAGCAGAGCTGCAGGAGCAGCTG 4863

415 ValSerLeuGluLysAspThrLeuGlnLeuAspLysGluMetGluLysAlaArgSer 434  
 4864 CGGAGCTGGAG-----CAGCTGCAGAGGAAACAAAGAG----- 4899  
 435 LeuSerLeuMetGluGluAlaAlaLeuAla-LysArgMetTyrAsnGlnGlnIleLysArg 454  
 4900 -----CTGCGAGCTGAAGCTGAACGGCTGGGCCCATGAGCTACAGCAGCGGTGGCTGAAG 4953  
 454 gProArgLeuSerProMetGluMetProValThrSerSerSerTyrSerProIleTyr 474  
 4954 ACCAAG-----AGCTGAACAGACCTGCCGCCACCTTACTGCCAGGTGCCGAGCTG 5007  
 474 rArgAspArgSerPheProSerGlnArgAspAspGlnAspGluLysSerAla----- 492  
 5008 GAGGCACAGG---TTGCCCATGCAGCAGCAGCTTCGAGACCTGGGCAATTCAGGTG 5064  
 493 -----LeuValSerSerTyrLeuGlyProSer----- 501  
 5065 GCACTGATGCTTTAAAGAGCCGTGAGCCCGAGGTAAAGCCCGAGCTGGACTTGAGTATT 5124  
 502 -ThrSerPheProHisArgSerArgSerProGluTyrMetValPro----- 517  
 5125 GACAGCTGATCTGAGCTGCGAGGAGGGGACCCCACTCAGTATCACCAGCAAGCTGCCT 5184  
 518 -LeuProHisGlyGlyLeuGlyArgSerValTyrAlaTyrGluHisLeuAlaProAsnSe 537  
 5185 CGTACCCAGCCAGCAGCGCAGCGCTCCCTGGAGAACCCAGCTCACCCTATCTCCAGCGC 5244  
 537 rTyrSerPro-GlyHisGlyHisArgLeuHisArgGlnTyrSerProSerLeuValHisG 557  
 5245 CTGCCCCCAAGGT-----AGAATCCCTGGAGAGTCTCTACTTCACT 5286  
 557 LysGlnArgHisProLeuGlnTyrSerProIleHisGlyGlnGlnLeuProTyrG 577  
 5287 CC-----CATCCCTCTCGAGTCAAGCCCGCCCTGGAGAGCAGCTCGACTCCCTGGG- 5339  
 577 LysLeuArgVal----- 581  
 5340 -----AGAGCTCTCTCGAGCTCGGTCTGTAAGACCGCTCGCTCGCGCAGCAGCAG 5394  
 582 ---TyrArgHisSerProSerGluLysArgTyrLeuGlyLeuSerAsnGlnArg 598  
 5395 CAGATCATCAATCACTACCATGACCAAGAGCTAGATGTGGAAGAGCCAGCAGCGC 5450

RESULT 14  
 US-08-195-487-3  
 ; Sequence 3, Application US/08195487  
 ; Patent No. 5783403  
 ; GENERAL INFORMATION:  
 ; APPLICANT: TOURKATLY, GARY  
 ; APPLICANT: LIDGARD, GRAHAM P  
 ; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
 ; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
 ; NUMBER OF SEQUENCES: 6  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: TESTA HURWITZ & THIBEAULT  
 ; STREET: 53 STATE STREET  
 ; CITY: BOSTON  
 ; STATE: MA  
 ; COUNTRY: USA  
 ; ZIP: 02109  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/195,487  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/901,701  
 ; FILING DATE:

415 ValSerLeuGluLysAspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSer 434  
4864 CGGAGCTTGAG-----CAGCTCAGAGGAAACAAGAG-----4899  
435 LeuSerLeuMetGluGluAlaLeuAla-LysArgMetTyrAsnGlnGluLysArg 454  
4900 -----CTGCGAGCTGAAGCTGAACGGCTGGCCATGAGCTACAGCAGCGTGGCTGGAAG 4953  
454 gProArgLeuSerProMetGluMetProProValThrSerSerSerTyrSerProLeuTyr 474  
4954 ACCAAG-----AGGTGACACAGCTCGCCCACTTACTGCCAGGTGGCGACCTG 5007  
474 rArgAspArgSerPheProSerGlnArgAspAspGlnAspGluLysSerAla-----492  
5008 GAGGCACAGG---TTGCCCATGACAGCAGCAGCAGCTTCGAGACCTGGGCAAAATTCAGGTG 5064  
493 -----LeuValSerSerTyrLeuGlyProSer-----501  
5065 GCAACTGATGCTTTAAGAGCGGTGAGCCCGAGCTTAAGCCCGAGCTGGACTTGAGTATT 5124  
502 -ThrSerPheProHisArgSerArgArgSerProGluTyrMetValPro-----517  
5125 GACAGCTGGATCTGAGTGGAGGAGGGAGCCCACTCAGTATCACCAGCAAGCTGCCT 5184  
518 -LeuProHisGlyGlyLeuGlyArgSerValTyrAlaTyrGluHisLeuAlaProAsnSe 537  
5185 CGTACCCAGCAGCAGCAGCAGCAGCTCCCTGGAGAACAGCCTCACCATTCTCCAGCGC 5244  
537 rTyrSerPro-GlyHisGlyHisArgLeuHisArgGlnTyrSerProSerLeuValHisG 557  
5245 CTGCCCCCAAGGT-----AGATCCCTGGAGAGTCTCTACTTCACT 5286  
557 lyGlnArgHisProLeuGlnTyrSerProProLleHisGlyGlnGlnGlnLeuProTyrG 577  
5287 CC-----CATCCCTGCTCGAGTCAAGCCCTGGAGAGCAGCTGGACTCCCTGGG- 5339  
577 lyLleGlnArgVal-----581  
5340 -----AGAGCTCTCTCGAGCTCGGTGTAAGCCCGCTCGCTCGTGGCGCAGCAG 5394  
582 -----TyrArgHisSerProSerGluGluArgTyrLeuGlyLeuSerAsnGlnArg 598  
5395 CAGATCATCAATCATTCAATGACCAAGAGCTAGATGTGAAGAGCCAGCAGCAGCGC 5450

RESULT 13  
US-08-467-781-3  
Sequence 3, Application US/08467781  
Patent No. 5780596  
GENERAL INFORMATION:  
APPLICANT: TOUTKATLY, GARY  
APPLICANT: LIDGARD, GRAHAM P  
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSES: TESTA, HURWITZ & THIBEAULT  
STREET: 125 HIGH STREET  
CITY: BOSTON  
STATE: MA  
COUNTRY: USA  
ZIP: 02110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,781  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: PITCHER ESQ, EDMUND R

REGISTRATION NUMBER: 27,829  
REFERENCE/DOCKET NUMBER: MTP-013  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 248-7000  
TELEFAX: (617) 248-7100  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6306 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..6306  
PUBLICATION INFORMATION:  
AUTHORS: COMPTON, DUANE A  
AUTHORS: SZILAK, ILYA  
AUTHORS: CLEVELAND, DON W  
TITLE: PRIMARY STRUCTURE OF NUMA, AN INTRANUCLEAR  
TITLE: PROTEIN THAT DEFINES A NOVEL PATHWAY FOR  
TITLE: SEGREGATION OF PROTEINS AT MITOSIS  
JOURNAL: J. Cell Biol.  
VOLUME: 116  
PAGES: 1395-1408  
DATE: 1992  
US-08-467-781-3  
Alignment Scores:  
Pred. No.: 0.059 Length: 6306  
Score: 127.00 Matches: 161  
Percent Similarity: 34.61% Conservative: 102  
Best Local Similarity: 21.18% Mismatches: 275  
Query Match: 4.09% Indels: 222  
DB: 1 Gaps: 35  
US-09-890-475-1 (1-609) x US-08-467-781-3 (1-6306)  
QY 9 AlaAlaGlnProThrThrThrAlaAsnProLeuLeuGlnArgHisGlnSerGlu-----26  
Db 3331 GCTGCTGCAGGACAGAGCCACAGAGCCCAAGCTGGAAGCAGCTGCGGCGCAGGTGAGC 3390  
QY 27 -----GlnArgArgArgGlu-----LeuProLysLysLeuVal 36  
Db 3391 AAGTGTGAACAGCAATGCCAAGAGCAGCAGCAGGCTGACAGCTTGGACGCACCTC 3450  
QY 37 GluThrGluSerThrSer-----MetAspIleThrIleGlyGln 49  
Db 3451 CAGGCTGAGCGGGCTCCCGGGTGTGAGCGGACAGTGTCTGTGAGACTCTGAGGCGCCAG 3510  
QY 50 -----SerLysGlnProGlnPheLeuLysSerIleAsp 60  
Db 3511 TTAGAGGAGAGAGCCCGCAGAGCTAGGGCAGCAGTCAAGTGTCTTACCTGCGCCCAACGG 3570  
QY 61 GluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAspAspLeuGln 80  
Db 3571 GAGTTGGCTGCTTCCCGCACCAGGTACAGACACACAGCAAGCTGAAGATGATGGAAG 3630  
QY 81 LysHisIle-----GluSerIleGluAsnAlaIleAspSerLysLeu 94  
Db 3631 GCCCAGGTGGCGCGCGCGCAGAGCTGAGAGGAAATAAGCTCATCAGCAGCTTG 3690  
QY 95 GluSerAsnGlyValValLeuAlaAlaArgAsnAsnPheHisGlnProMetLeuSer 114  
Db 3691 GAGGAGGAGGTGTCTCATCTCTG-----AATCGCCAGGTCTCTGGAG 3729  
QY 115 ProProArgAsnAsnValSerValGluThrThrValThrVal-----SerGlnProSerGln 133  
Db 3730 AAGGAGGGGGAGAGCAGAGAGGTGAAGCGGCTGTGTGATGCGCGAGTGCAGAGAGAGCCAG 3789  
QY 134 GluIle-----ValProGluThr 139  
Db 3790 AAGCTGGAGGAGAGCTGCGCTGTGTGAGGAGCAGCAGCAGCAGCAGCAGCTG 3849

NAME: PITCHER ESQ, EDMUND R  
 REGISTRATION NUMBER: 27,829  
 REFERENCE/POCKET NUMBER: MTP-013  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 248-7000  
 TELEFAX: (617) 248-7100  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 6306 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..6306  
 PUBLICATION INFORMATION:  
 AUTHORS: COMPTON, DUANE A  
 AUTHORS: SZILAK, ILYA  
 AUTHORS: CLEVELAND, DON W  
 TITLE: PRIMARY STRUCTURE OF NUMA, AN INTRANUCLEAR  
 TITLE: PROTEIN THAT DEFINES A NOVEL PATHWAY FOR  
 TITLE: SEGREGATION OF PROTEINS AT MITOSIS  
 JOURNAL: J. Cell Biol.  
 VOLUME: 116  
 PAGES: 1385-1408  
 DATE: 1992  
 US-08-470-950-3

Alignment Scores:  
 Pred. No.: 0.059 Length: 6306  
 Score: 127.00 Matches: 161  
 Percent Similarity: 34.61% Conservative: 102  
 Best Local Similarity: 21.18% Mismatches: 275  
 Query Match: 4.09% Indels: 222  
 Gaps: 35

US-09-890-475-1 (1-609) x US-08-470-950-3 (1-6306)

```

;
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/POCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6306 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..6306
; PUBLICATION INFORMATION:
; AUTHORS: COMPTON, DUANE A
; AUTHORS: SZILAK, ILYA
; AUTHORS: CLEVELAND, DON W
; TITLE: PRIMARY STRUCTURE OF NUMA, AN INTRANUCLEAR
; TITLE: PROTEIN THAT DEFINES A NOVEL PATHWAY FOR
; TITLE: SEGREGATION OF PROTEINS AT MITOSIS
; JOURNAL: J. Cell Biol.
; VOLUME: 116
; PAGES: 1385-1408
; DATE: 1992
; US-08-470-950-3

;
; Alignment Scores:
; Pred. No.: 0.059 Length: 6306
; Score: 127.00 Matches: 161
; Percent Similarity: 34.61% Conservative: 102
; Best Local Similarity: 21.18% Mismatches: 275
; Query Match: 4.09% Indels: 222
; Gaps: 35

;
; US-09-890-475-1 (1-609) x US-08-470-950-3 (1-6306)
;
; QY 9 AlaAlaGlnProThrThrAlaAsnProLeuLeuGlnArgHisGlnSerGlu----- 26
; DB 3331 CTTCTGGCAGGACAGAGCCAAAGCCGCCAGCTGGAGCACTCGGGGACAGGTGAGC 3390
;
; QY 27 -----GlnArgArgGlu-----LeuProLysIleVal 36
; DB 3391 AAGCTGGAACAGCAATCCCAAGACAGCAGGAGCAGGCTGACAGCTGGAACAGGCTC 3450
;
; QY 37 GluThrGluSerThr-----MetAspIleThrIleGlyGln 49
; DB 3451 GAGCTGAGCGGGCTCCCGGCTGAGCGGACAGTGTCTTGAGACTCTGAGAGCTGCGG 3510
;
; QY 50 -----SerLysGlnProGlnPheLeuLysSerIleAsp 60
; DB 3511 TTAGAGGAGAGGCCAGGAGCTAGGGCAGTCAAGTGCCTTAGCCCTCGGCCCAACGG 3570
;
; QY 61 GluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAspAspLeuGln 80
; DB 3571 GAGTTGGTGGCTTCGACCAAGGTACAGACCAAGAGGCTGAAGATGAGTGGAG 3630
;
; QY 81 LysHisIle-----GluSerIleGluAsnAlaIleAspSerLysLeu 94
; DB 3631 GCCCAGGTGGCCCGCGCCGAGAGGTGAGAGGAAATAAGCTCATCAGAGCTTG 3690
;
; QY 95 GluSerAsnGlyValValLeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSer 114
; DB 3691 GAGGAGGAGGTGTCCATCTG-----AATCGCCAGGTCTCTGGAG 3729
;
; QY 115 ProProArgAsnAsnValSerValGluThrThrValThrVal-----SerGlnProSerGln 133
; DB 3730 AAGGAGGGGAGACAGAGGAGTTGAAGCGGCTGTGATGGCCGCTCAGACAGAGCCAG 3789
;
; QY 134 GluIle-----ValProGluThr 139
; DB -----ValProGluThr 139

```



Db 2877 CGAACACTCTGTACAATACTTACGAAGGTATCGATTGGCCGACTCTTCAGGAAG 2936  
Qy 394 MetGluThrHisLeuAspProAlaLysGluLeuProGlyTrpGlnIleLysGluGln 413  
Db 2937 CTGGAGCTGCAGCTGCTGAGTGGAGCGTGCAGCGGATGCTCGGAGTGAAGGAGCGT 2996  
Qy 414 IleValSerLeu-----GluLysAspThrLeuGlnLeu 424  
Db 2997 GTTCTGTCCAATGCCGTGCCAGCTAGCTGGAAGCGCAAGGAATCGGTGCTG 3056  
Qy 425 AspLysGluMetGluGluLysAlaArgSerLeuSerLeuMet-----Glu 439  
Db 3057 CAGCGCCAGCAAGTAA-----CTCACTTGTGCTGCTGCATGGCGGAG 3107  
Qy 440 GluAlaLeuAlaLysArgMetTyrSerGlnGlnIleLysArgProArgLeuSerPro 459  
Db 3108 GAAAGCTCGCCAGTTCGCCACGCTACGAGCAAGCAAGCTTTGGCCCT-----GCC 3158  
Qy 460 MetGluMetProProValThrSerSer-----TyrSerProIleTyrArgAsp 476  
Db 3159 TGTGTGCCCGCCCGTGCAGCCAGTAGTGTGACTAGCATCTTCTACAAGGT 3218  
Qy 477 ArgSerPheProSerGlnArgAspAspGlnAspGluIleSerAlaLeuValSerSer 496  
Db 3219 GCGCGCAATCCAAAGCAAG-----GCACGTATC----- 3245  
Qy 497 TyrLeuGlyProSerThrSerPheProHisArgSerArgArgSerProGluTyrMetVal 516  
Db 3246 -----AAAGGCCAAG----- 3257  
Qy 517 ProLeuProHisGlyGlyLeu-----GlyArg 525  
Db 3258 CCGATCCCGCCGCTCGCGCTAGACAATAACAAGACAAAGACGAGCGGACGC 3317  
Qy 526 SerValTyrAlaTyrGluHisLeu-----AlaProAsnSerTyrSer 539  
Db 3318 AACATCTTGGCAAGTTCATGTGCTCACCAGATGCTGCTGCGGAGCGGTTC 3377  
Qy 540 ProGly-HisGlyHisArgLeuHisArgGlnTyrSerProSerLeuVal-----HisG 557  
Db 3378 TCAGGTCCAGCGCCGAGCAATCCAC-----GGCCCTGATTGAGTCAACCACACG 3428  
Qy 557 LysArgHisProLeuGlnTyrSerProProIle 568  
Db 3429 GCCA-CGGCAACCACTACGTGACACACCACTG 3462

RESULT 11

US-08-466-390-3  
Sequence 3, Application US/08466390  
Patent No. 568562  
GENERAL INFORMATION:  
APPLICANT: TOKUATLY, GARY  
TITLE OF INVENTION: LIDGARD, GRAHAM P  
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
STREET: 125 HIGH STREET  
CITY: BOSTON  
STATE: MA  
COUNTRY: USA  
ZIP: 02110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/466,390  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: PITCHER ESQ. EDMUND R  
REGISTRATION NUMBER: 27,829  
REFERENCE/DOCKET NUMBER: MTP-013  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 248-7000  
TELEFAX: (617) 248-7100  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6306 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..6306  
PUBLICATION INFORMATION:  
AUTHORS: COMPTON, DUANE A  
AUTHORS: SZILAK, ILYA  
AUTHORS: CLEVELAND, DON W  
TITLE: PRIMARY STRUCTURE OF NUMA, AN INTRANUCLEAR  
TITLE: PROTEIN THAT DEFINES A NOVEL PATHWAY FOR  
TITLE: SEGREGATION OF PROTEINS AT MITOSIS  
JOURNAL: J. Cell Biol.  
VOLUME: 116  
PAGES: 1395-1408  
DATE: 1992  
US-08-466-390-3  
Alignment Scores:  
Pred. No.: 0.059 Length: 6306  
Score: 127.00 Matches: 161  
Percent Similarity: 34.61% Conservative: 102  
Best Local Similarity: 21.18% Mismatches: 275  
Query Match: 4.09% Indels: 222  
DB: 1 Gaps: 35  
US-09-890-475-1 (1-609) x US-08-466-390-3 (1-6306)  
Qy 9 AlaAlaGlnProThrThrAlaAsnProLeuGlnArgHisGlnSerGlu----- 26  
Db 3331 GCTGCTGGCAGCAGACAGCCCAACAGAGCCCAAGCTGGAGCACTGCGGGCAGAGTGAGC 3390  
Qy 27 -----GlnArgArgGlu-----LeuProLysIleVal 36  
Db 3391 AGCTGGAACAGCAATGCCAAGCAGCAGCAGCTGACAGCTGACAGCTGGACGCGCTC 3450  
Qy 37 GluThrGluSerThrSer-----MetAspIleThrIleGlyGln 49  
Db 3451 GAGGCTGAGCGGGCTCCCGGGCTGAGCGGACAGTCTCTGGAGACTCTGCGAGGCGCAG 3510  
Qy 50 -----SerLysGlnProGlnPheLeuLysSerIleAsp 60  
Db 3511 TTAGAGGAGAGCGCCAGGAGCTAGGCGACAGTCAAGTGCCTTAGCTCGGCCAACGG 3570  
Qy 61 GluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAspLeuGln 80  
Db 3571 GAGTTGGCTGCTTCCGCAACCAAGCTACAGCACCAAGCAAGCTGAAGATGATGAGTGAAG 3630  
Qy 81 LysHisIle-----GluSerIleGluAsnAlaIleAspSerLysLeu 94  
Db 3631 GCCCAGGTGGCCCGCGCGGCGGCAAGAGCTGAGAGGAAATAATAGCCCTCATCAGCAGCTTG 3690  
Qy 95 GluSerAsnGlyValValIleAlaAlaArgAsnAsnPheHisGlnProMetLeuSer 114  
Db 3691 GAGGAGGAGGTGCTCCATCCTG-----AATCGCAGGCTCTGGAG 3729  
Qy 115 ProProArgAsnValSerValGluThrThrValThrVal-----SerGlnProSerGln 133  
Db 3730 AAGGAGGGGAGAGCAGAGAGTGAAGCGGTGGTGTGATGCGCGGATCAGAGAGCAG 3789  
Qy 134 GluIle-----ValProGluThr 139

```

1260 ATCAAGTTCAGAAATCCAGCCCTGCTACATCTCTCTTTGTAGAACTTGATGAAGAAG 1319
1410 ILeuys
1320 GTGAAGACGACCAACTGAACCTACTGTAAGTGGAGAGGTGTCTCCCGCACTGGAACCC 1379
421 ThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSerLeuMetGluGlu 440
1380 ACAGAAACGCTGCTGAGTGAAGAGGATGAACCGCAAGGAGAGAGCCCTTGTGAAGAG 1439
441 AlaAlaLeuAlaLysArgMetTyAsnGlnGlnIleLysArgPro----- 455
1440 CTGTCCTCCCTGCCAGCGAG-----AAGAAGCCCTGCGCGCTCTGTGAG 1481
456 -----ArgLeuSerProMet---GluMetProValThrSerSer 468
1482 GGCAAGCTTAGACTGTCCCCCGCGGTGAGTGAAGCCCATGCCGTGCT 1532
RESULT 10.
JS-09-098-901-1
Sequence 1, Application US/09098901B
Patent No. 6218144
GENERAL INFORMATION:
APPLICANT: Scott, Matthew
APPLICANT: Sisson, John C.
TITLE OF INVENTION: Costal2 Genes and their Uses
FILE REFERENCE: SUN-65P
CURRENT APPLICATION NUMBER: US/09/098,901B
EARLIER FILING DATE: 1998-06-17
EARLIER APPLICATION NUMBER: 60/051,347
EARLIER FILING DATE: 1997-06-30
NUMBER OF SEQ ID NOS: 11
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 3605
TYPE: DNA
ORGANISM: D. Melanogaster
JS-09-098-901-1

Alignment Scores:
Pred. No.: 0.0207 Length: 3605
Score: 127.50 Matches: 150
Percent Similarity: 35.48% Conservative: 103
Best Local Similarity: 21.04% Mismatches: 244
Query Match: 4.11% Indels: 218
DE: 3 Gaps: 34

JS-09-890-475-1 (1-609) x US-09-098-901-1 (1-3605)
2Y 18 ProLeuLeuGlnArgHisGlnSerGluGlnArgArgGluLeuProLysIleValGlu 37
1492 CCCATTCTGGAGCGCGAGGAGCCCGAGACCGTGAATCTCGAGAGCAATTCGGAGTCG 1551
38 ThrGluSerThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLys 57
1552 CCAAACTCCGACACGAAACGACACACACACATGAGTCGCGATCGGCCGATCGGACGAC 1611
58 SerIleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAsp 77
1612 AAGTAGAAGATCTA-----ATGGAAGAGTTTCGCGACAAACAGAC 1653
78 Asp-----LeuGlnLysHisIleGluSerIleGluAsn 88
1654 GCTCTTATCTATGAAAAACACGCTGAATATCTATCCAAAGCATCCGAAGCGGTTATGCAA 1713
89 AlaIleAspSerLysLeuGluSer-----ArgLys-----Val 99
1714 AGCCAAAGACCGGAGATTGAGGCGACGCGCGACGAAAGAAATGTTGATGATCGAAAGTC 1773
100 ValLeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSerProProArgAsnAsn 119
1774 AGCATTTGCAGCTCGGAGGAGAGTGT---CAGCCAGGTGCTAGCTTAAGTACTGCTGAG 1830

```

```

120 ValSerValGluThrThrValThrValSerGlnProSerGlnGluIleValProGluThr 139
1831 CTTGCCATGCTTAATCGGGTAGCTTCCAGCAGCGCGCTCCGCCCATGATCCTGAGTCG 1890
140 -----SerAsnLysProGluGlyArgMetCysGluLeuMetCys 153
1891 GTCGTCGATCTCTGTGAAAGTTCTTCGGCGGAGGAATCCGTCAGCGCGCTCTCGCTGCC 1950
154 Ser-----LysGlyLeuArgLysTyIleTyAla 163
1951 GCGCGCGCATCTCTCTATTGAACAGCTGCAGAAAAAATTCGCGCAACTGGTC---GCT 2007
164 AsnIleSerAspGlnAlaLysLeuMetGluGluIle----- 175
2008 GAGATCGAGGCGAAACACACGACAGTTACGGGAAATCGAAGAACATCCAGGTAAACAAA 2067
176 -----ProSerAlaLeuLys-----Leu 181
2068 ATATAATCGCCGAATTG-GTCAAGAACACGCGATACAGCAGCATGCAAGCAAGATTT 2126
182 AlalysGluProAlalysPheValLeuAspCys-----IleGlyLys 195
2127 CACAGAAACGTCCTCAAACTTGAGCGCGAGTGCACAGGCCAAGAGCAGTTAGTAAAG 2186
196 PheTyLeuGlnGlyArg----- 201
2187 GCCTAGTTCAGGCGGGGTGTCGCGAGATTGAGCATGACCAACCATATTCGGACAT 2246
202 -----ArgAlaPheThrLysGluSerProMet-----SerSerAlaArg 214
2247 CTCGAGCGTCGACTAGACACCTCAGCTCAATGAAGCATATTGCGGTGAGAGCGGACAG 2306
215 GlnValSerLeuLeu-----IleLeuGluSerPheLeuLeuMetProAspArgGly 231
2307 AAGGTGAAGAGCTACAGCAATCGTGGCGGAGTCCGAAACACAGCCCGATATTACAG 2366
232 LysGlyLysValLysIleGluSerTrpIleLysAspGluAlaGluThrAlaAlaValAla 251
2367 AAA---AAGTTCGAAGAGGTGCAAGCTGCGCTGCCATGAGCGCGGAGTGTGCTCAA 2423
252 TrpArgLysArgLeuMetThrGluGlyLeuAlaAlaGlu-----LysMet 268
2424 CTACGAGAACCCAGGAGACTGGCAAGAGTAGTAGAGCGCAAGGTTCTCCCGAGCAA 2483
269 AspAlaArgGlyLeuLeuValAlaCysPheGlyValProSerAsnPheArgSer 288
2484 CAGCGCCGCAAGTTAAGCGAGTACAGCT-----AGGATT 2519
289 ThrAspLeuLeuAspLeuIleArgMetSerGlySerAsn---GluIleAlaGlyAlaLeu 307
2520 ACGCACCTTAATCACATTTTACGCGAGAGTTCGATAACCTGAGAGAGCAGCGGAGCCA 2579
308 LysArgSerGlnPheLeu-----ValProMetValSerGlyIleValGluSerSer 324
2580 GAACGAGGAGACCTTGCCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2639
325 IleLysArgGlyMetHis----- 330
2640 TTGAAGACGCTGCTCATTTGGACCGCAAACTTAAGCGGACAAAGGTGCTGACGCAAAAG 2699
331 -----IleGluAlaLeuMetValTyThr 339
2700 GAGGAGCGCAAGCTGCTCGAGTCGATGAGCGCATGAGCGCATGAGCGCATGAGCGCATGAA 2759
340 PheGlyMetGluAspLysPheSerAlaAlaLeuValLeuThrSerPheLeuLysMetSer 359
2760 TTCAAGAACGAGATGATCAGCGGCCCGCTCCATCGACACGAGCGAC---CGAATTCAG 2816
360 LysGluSerPheGluArgAlaLysArgLysAlaGlnSerProLeuAlaPheLysGluAla 379
2817 CGGAGAGGAGGAGAACACAGATGCTGAGCGCGCTAAATGCTCTCTCAACGAGGAGAGTG 2876
380 AlaThrLysGlnLeuAlaValLeuSerSerValMetGlnCys----- 393

```



```

NAME/KEY: CDS
LOCATION: (118)...(3012)
US-09-620-312D-372

Alignment Scores:
red. NO.: 0.0341
core: 130.50
Percent Similarity: 30.80%
Best Local Similarity: 18.92%
Query Match: 4.20%
JB: 4

US-09-890-475-1 (1-609) x US-09-620-312D-372 (1-7393)

Y 1 MetSerAsnTyrPro-----ProThrValAlaAlaGlnPro 12
b 859 ATGCCCAAGAGTCCCAACTCAAGCCCATTCAGCCCAAGCCCACTGTTATGGAGAACCT 918
Y 13 ThrThrThrAlaAsnProLeu-----GlnArgHisGlnSerGluGlnArg 28
b 919 ---TTCAACAGTCAACCTGGCTTGACTCCAGCCCAAGCAAGCAAGAAAGAAAGCAAA 975
Y 29 ArgArgGluLeuProLysIleValGluThrGluSerThrSerMetAspIleThrIleGly 48
b 976 AAGAAGGAATCTTCAAGGAACCTTGAAGTCTCTGACCCCTGGGAAGGTGTGTGAGCA 1035
Y 49 GlnSerLysGlnProGlnPheLeuLysSerIleAspGluLeuAlaPheSerValAla 68
b 1036 GAGGAAGCAAAAGCCCATTC-----1056
Y 69 ValGluThrPheLysArgGlnPheAspAspLeuGlnLysHisIleGluSerIleGluAsn 88
b 1057 -----AGGGAATCTTCAGAAATCGGATGAAATGGAGGGCTCTTAAT 1101
Y 89 AlaIleAspSerLysLeuGluSerAsnGlyValValLeuAlaAlaArgHisAsnAsn 108
b 1102 GGCTCATCAGACCCCAAGCCGACTGGCTAGCATCAAGCTGAAGCCGCAAGATC 1161
Y 109 HisGlnProMetLeuSerProArgAsnValSerValGluThrThrValThrVal 128
b 1162 TACAGTTTCAGGCAATGCCCCAGCCCTTCATGGAGGAGTAGCCGCTTGAAC 1221
Y 129 SerGlnProSerGlnGluIleValPro-----GluThrSerAsnLysProGluGly 145
b 1222 ACTACCCCTACTCAGCCCTGACTCCCTTACATGGTGACCCAGATGAGCTGAGCC 1281
Y 146 GlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyrIleTyrAlaAsn 165
b 1282 AGC-----TCAGTCAAAACCAACAGCCCTGCTGATCTGACATC 1320
Y 166 SerAspGlnAlaLys-----LeuMetGluGluIleProSer-----AlaLeu 179
b 1321 TCTGATGCTGGGAGGATGGGGGCAAGGTAGACAGTGTCAAATCAAAGGACGCCGAA 1380
Y 180 LysLeuAlaLysGluProAlaLysPheValLeu-----AspCysIleGly 194
b 1381 CAGTTGGTTAAAGAGGGGCTAAGAAACTTTTTCCTCCCTCAGCCTCAGACCAAGAC 1440
Y 195 LysPheTyrLeuGlnGlyArgAlaPheThrLysGluSer-----208
b 1441 TCACCATATTACCAAGGTTTGAGGTACTATTCTCCAGTTATGACAGTCCAGCCCT 1500
Y 209 -----PrometSerSerAlaArgGlnValSerLeuLeuLeuLeuLeuLeuSerPhe 224
b 1501 GGGGCTCTGAACCCAGCAGCCAGCAGGA-----GTGAGAGCCAG 1542
Y 225 LeuLeuMetProAspArgGlyLysLysValLysIleGluSerTrpIleLysAspGlu 244
b 1543 GCCCTGAGACAAAGAGGATGAGGAACCTGAGCATAGAGGAAGTGAAGAACGAT 1602
Y 245 AlaGluThrAlaAlaValAlaTrpArgLysArgLeuMetThrGluGlyGlyLeuAla 264
b 1602 -----1602

265 AlaGluLysMetAspAlaArgGlyLeuLeuLeuValAlaCysPheGlyValProSer 284
1603 -----ATCTGTGAAGAAAAGAGGCC 1623
285 AsnPheArgSerThrAspLeuLeuAspLeuIleArgMetSerGlySerAsnGluIleAla 304
1624 GAGCTGAGCAGTTCAGTCCAGAGCCCTCGGTATCCAGCAGCGTCCCAATATGTACATG 1683
305 GlyAlaLeuLysArgSerGlnPheLeuValProMetValSerGlyIleValGluSerSer 324
1684 CAGTCCCTGTACTACCAACCACTGCTATGTATGATGCTATGCTATGCTATGCTATGCT 1743
325 IleLysArgGlyMetHisIleGluAlaLeuGluMetValTyrThrPheGlyMetGluAsp 344
1744 TACCAC-----ACCCACCTTCTGAGCACTAACAGGCTTACCGCAGCAGTACCAAGAA 1797
345 LysPheSerAlaAlaLeuValLeuThrSerPheLeuLysMetSerLysGluSerPheGlu 364
1798 -----CAGCAGAAACGCCAGAGCTTACAG 1821
365 ArgAlaLysArg-----LysAlaGlnSerProLeuAlaPheLysGluAlaAla 380
1822 CAGCAGCAGCGGGAGTGGACCAAGAGGAGGAGATGGGCTGAAGGAGCGGAGGAGCA 1881
381 ThrLysGln-----LeuAlaValLeuSerSer 389
1882 CTCAGGAAGAGTGGAAAGCAAGCCGCTCAATTCACCACTCTCACCAAGCCCCAGC 1941
390 ValMetGlnCysMetGluThr-----396
1942 CTGACAGACCTGTGAAATCAGGACCTGGCAAGCCAGAGCCAGGGCTGACCCAGCC 2001
397 -----HisLysLeuAspProAlaLysGluLeuProGlyTrpGlnIleLys 411
2002 AATCAGTCTATCTCCAGATTAGATGACTCTTCAAACTCCCGGC-----CAGCCCT 2058
412 GluGlnIleValSerLeuGluLysAspThrLeuGlnLeuAspLysGluMetGluGluLys 431
2059 GAAGCCCTTAAAGTGAAGCTGAGTGATGCCAGCACCTAAGCAAGAGGAGCTCTGAGGC 2118
432 AlaArgSerLeuSerLeuMetGluGluAlaAlaLeuAlaLysArgMetTyrAsnGln 451
2119 AAGCAGGCTGCTGAGTGTGTCAGCAGGAGATGGATGATGATGATGATGATGATGATG 2178
452 IleLysArgProArgLeuSerProMetGluMetProProValThrSerSerTyrSer 471
2179 GAGCAGAGCCCGGATGTGGACATATGTTATCT-----GCCAAGTACTCA 2226
472 ProIle-----TyrArgAspArgSerPheProSerGln 482
2227 GACATCAAGTCAGAGGATGAGCGTGGAAAGGAGGAGCGGAGCGCAAAATTGAAGGAG 2286
483 ArgAspAspAspGlnAspGluIle-----490
2287 AGGAGTGGAGTAGGACTCTGTCCCAAGGAAGATGGGAGGAAGACACAAGTAGTGAC 2346
491 -----SerAlaLeuValSerSerTyrLeuGlyProSerThr 502
2347 TGCAGCTGCCCACTGTCAGAGGAGTCTCGCTTGGGAGCAAGAGCCCGGCAAGTGTG 2406
503 SerPhePro-----HisArgSerArgSerProGluTyrMet 515
2407 CATGTGCTGTGCTCTCCCTCCCTTACCAGCAGCAGTCC-----TAC 2448
516 ValProLeuProHisGlyGlyLeuGlyArgSerValTyrAlaTyrGluHisLeuAlaPro 535
2449 ATCCCTTACATGCACGCG-----TATTCCTACAGT-----2478
536 AsnSerTyrSerProGlyHis-----GlyHisArgLeuHisArgGlnTyrSerProSerLeu 554
2479 CAGTCTTACAGACCCCAACCAACCCAGCTACCG-----AGCATGCTGCT 2523

```

QY 226 -----LeuMetProAspArgGly-----LysGlyLysVallys 236  
Db 1203 ATCCCATCTGCTCACTCCACTGCTTGACAAAGAGAGGAGCGCGGAAGCTTCC 1262  
QY 237 IleGluSer-----Trr----- 240  
Db 1263 ATCTCACCACCTCTCTCGAGCGCTCTGGAAGTGTGACCGTCCGATGCTCTCCACCC 1322  
QY 241 ---IleLysAspGluAlaGluThrAla-----AlaValAlaTrr 252  
Db 1323 GTTGTCTGATGAGCAGACACTGCTCTGGGTCTGGTCTACAGCAGTCCGAGAGTGTG 1382  
QY 253 ArgLysArgLeuMetThrGluGlyLysLeuAlaAlaGluLysMetAspAlaArgGly 272  
Db 1383 AACGAGCCCTCAGGACACAGAGTGTCTATCAGAGCGGAAG-----CGCGGT 1433  
QY 273 LeuLeuLeuValAlaCysPheGlyValProSerAsnPheArgSerThrAspLeuLeu 292  
Db 1434 CTCTGGTACAGGGTGTCTACTCCGA-----GACACTCAG 1469  
QY 293 AspleuileArgMetSerGlySer----- 300  
Db 1470 GAGCTGTGCGCATCTCGTGTACTCGATACGATGCTCTCAAGTTGTCTGTGAAGCAG 1529  
QY 301 -----AsnGluLeuAlaGlyAlaLeu 307  
Db 1530 AAGGGTCTGTTGCCACCTCAGTCAGTCGGTCTGTTGGTCTGAGTCTCAAGGCTTCC 1589  
QY 308 LysArgSerGlnPheLeuValProMetValSerGlyIleValGluSerSer----- 324  
Db 1590 AAGCTCAGCAGACTTGTATTTCGAGGAAACAGTCTGCCCGGAGGCTCTTACACTGCC 1649  
QY 325 -----IleLysArgGlyMetHisIleGluAlaLeuMetVal 337  
Db 1650 CGTCTCTTCCGATGAGAGTAGTCCCGGCCAAGATCATGAGGAGGCTGAGGAGCTC 1709  
QY 338 TyrThr-----PheGlyMetGluAspLysPheSerAlaAla 349  
Db 1710 TGACCCCTCAGACCCCGAGAAATCGCTTTGAGGCTCGCATCTCTTCTACTTGTCT 1769  
QY 350 LeuValLeuThrSerPheLeuLysMetSerLysGluSerPheGluArgAla----- 366  
Db 1770 CTTACCAAGGCGGTGTGTCGCGGTTACTCTTCCGATATCGAAAGGAGCCTTGACGCC 1829  
QY 366 ----- 366  
Db 1830 AAGAGCTGGAAGTCAAGCGCAGGACTGAGATGCTAAGGGTAAGTGGCTGAGAAGGAG 1889  
QY 367 -----LysArgLysAlaGlnSerProLeuAlaPheLysGluAlaAlaThrLysGlnLeu 384  
Db 1890 GGATCAAGCGTGGCGGTCCGCTCCGCTGCCACTTCCGCGCTGTCCACCAAGGAGGCC 1949  
QY 385 AlaValLeuSerSerValMetGlnCys-----MetGlu 395  
Db 1950 GCCCAGAGACCCCTCGAAGATCACCATGAGAGCTTCCAGCGCTCCAGGCTCTCT 2009  
QY 396 ThrHisLysLeuAspProAlaLysGluLeuProGlyTrrPdnIleLysGluGlnIleVal 415  
Db 2010 ACCGAGGAGCTGATGCTGCTCTCAAGCGTCTCGCGCAAGAGTCCGATGCTATCTAC 2069  
QY 416 SerLeu-----GluLysAspThrLeuGln 423  
Db 2070 AAGATCATTTGCCCATCATCATGAGAGCTCCGCAAGACCGGCGAGAGCTTCTGTCTG 2129  
QY 424 LeuAspLysGluMetGluGluLysAlaArgSerLeuSerLeuMetGluGluAlaLeu 443  
Db 2130 TACACTCAAGTTCGAGAGGAGCTCTCTTACTAGCCC-CGTCTCGAAGCGCGCTT 2188  
QY 444 AlaLysArgMetTyrAsnGlnGlnIleLysArgProArgLeuSerProMetGluMetPro 463  
Db 2189 CCCAAGGAGGTTATGCG-AGTCCCTGAGGAGACCA---TGTGTCATCAGCTGTCTCT 2244  
QY 464 ProValThrSerSerSerTyrSerProIleTyrArgAspArgSerPheProSerGln-Ar 483

Db 2245 TCGAACAATCCGCAAGTTCCACGCCGCCAGAGGAGGAGAAC---CCTCCAGGTGG 2301  
QY 483 GAspAspAspGlnAspGlnIleSerAlaLeuValSerSerTyrLeuGlyProSerThrSe 503  
Db 2302 AGACATCCCGGTGTCTCTGACGCGTTCTCTGTCCTCCATCGAGCGCTGGCTGCT 2361  
QY 503 rPheProHisArgSerArgArgProGluTyrMetValProLeuProHisGlyGlyLe 523  
Db 2362 ACATCCCGGGGTACTCCGCGTCTCTCCCGACAC-----TGCCCTTATGCTGGGTG 2412  
QY 523 uGlyArgSerValTyrAlaTyrGluHisLeuAlaProAsnSerTyrSerProGlyHisG1 543  
Db 2413 TTCCCGC-----CATGG 2424  
QY 543 yHisArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGlnArgHisProLeuG1 563  
Db 2425 TCGCGGCTG-----CAACAAGATTGTGTTCG 2451  
QY 563 nTyrSerProPro-----IleHisGlyGlnGlnLeuPro----- 575  
Db 2452 CTTCTCTCCCGCGCCAGCAACCATCATCTCCGAGATTGTCCAGTCCGCTCACAAGG 2511  
QY 576 -TyrGly-----IleGlnArgValTyrArgHisSerProSer----- 587  
Db 2512 TTGGGCGCGAGTCCATCTGCTTGGCGCGGTGCCAGCGCTAGCTGCCATGCGCTAAG 2571  
QY 588 ----GluGluArgTyrLeuGlyLeuSerAsnGlnArgSerProArgSerAsnSerSerLe 606  
Db 2572 GCACCGAGAGCATCAACCAAGTCGCAAGATTCTCGGCC-CCGTAACCAAGTTCGTCACT 2630  
QY 606 uAspProLys 609  
Db 2631 GCTGCCAAGA 2640

RESULT 8  
US-09-620-312D-372  
; Sequence 372, Application US/09620312D  
; Patent No. 6569662  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Chen, Rui-kong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yungling  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: John Tillinghaast  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; CURRENT FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: pt\_FL\_genes Version 1.0  
; SEQ ID NO 372  
; LENGTH: 7393  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:

```

10345 AGGAGCTGGCCCTCGCTTACACGCGGACCTCTCCGAGGTGGCGGAGATGGCGGAGGCG- 10403
261 lyLeuAlaAlaAlaGluMetAspAlaArgGlyLeuLeuLeuValAlaAlaCysPheG 281
10404 -----ATGGGTGCCGAACCTTGACCTAGCCAGCGCTGGCGGAACCTCACTCCG----- 10448
281 lyValProSerAsnPheArgSerThrAspLeuLeuAspLeuLeuLeuArgMetSerGlySerA 301
10449 -----GAATCCGGCGACCTGCTGGAG----- 10469
301 snGluIleAlaGlyAlaLeuLeuArgSerGlnPheLeuValProMetValSerGlyIleV 321
10470 -----CAGGAAGATGACGCGCGCATCATCCGCTGATCAACGCGCATCC 10512
321 alGluSerSerIleLeuArgGly-----MetHisIleGluAlaLeuGlu-----M 336
10513 TCACGAGGGGATCAAGCGCGCGCTCCGACATCCACCTGGAAACCTTCGAAACGCG 10572
336 etValTyrThrPheGlyMetGluAspLysPheSerAlaAlaLeuValLeuThrSerPheL 356
10573 TGGTGGTGGCTTTCGCTGACGCG-----ATCCTCCGCGAAGTGA 10614
356 euLysMetSerLysGluSerPheGluArgAlaLysArgLysAlaGlnSerProLeuAlaP 376
10615 TCGAACCGCGCGCGGAG-----CTGGCG- 10637
376 heLysGluAlaAlaThrLysGlnLeuAlaValLeuSerSerValMetGlnCysMetGluT 396
10638 -----CGCTGCTGCTTCGCGGTCGCGGTCGCGGTCGCGG- 10670
396 hrHisLysLeuAspProAla---LysGluLeuPro----- 406
10671 -----CGCTGGACATCCGCGAGAGCGCGTACCGCAGGAGCGCGTATTTCGCTCAAGG 10725
407 -----GlyTrpGlnIleLysGluIleValSerLeu-----GluLysA 420
10726 TCGCGGTTCGGAGTGATATCGGTCTCCACCTCGCTCGCGCCACGCGGAGCGGG 10785
420 spThrLeuGlnLeuAspLysGluMetGluLysAlaArgSerLeuSerLeu----- 437
10786 TGGTGTGCTGCTG-----CTGCACAAGCAGCGCGCGCTGCTGCTCACGCATC 10836
438 -----MetGluGluAlaAlaLeuAlaLysArgMetTyrAsnGlnGlnIleLysArgPro- 455
10837 TGGGCATGAGCAGCGCG-----GACCGCGCTGCTCGACGACACACCTGCGCAAGCGCG 10890
456 -ArgLeuSerProMetGluMetProValThrSerSerSerTyrSerProIleTyrAr 475
10891 ACGCATCATCTAGTCACCGGCCCCCAGCGCTCGGGCAAGACCACACCTGTACGCG 10950
475 gAspArgSerPheProSerGlnArgAspAspGlnAspGluIleSerAlaLeuValse 495
10951 GCCT---GGTCAACCTCAACGA----- 10969
495 rSerTyrLeuGlyProSerThrSerPheProHisArgSerArgArgSerProGluTyrWe 515
10970 -----CCGTCGCGCAATATCTCATCGTGAAGA-----CCGATCGAGT 11010
515 tValPro-----LeuProHisGlyGlyLeuGlyArgSerVa 527
11011 ACTACCTGGAGGATCGCGCAGACCCAGGTCAACCGCGGTGGACATGACCTTCGCC 11070
527 lTyrAlaTyrGluHisLeuAlaPro-----AsnSerTyrSerProG1 541
11071 CGGCGCTCGCGCATCTCTCGCGCAGACCCGACGCTGTGTGTGTCGGCGAGATCCGCG 11130
541 yHisGlyHisArgLeuHisArgGlnTyrSer-----ProSerLeuVal-- 555
11131 ACCAGGAGACCGCGCATCGCGGTGAGCGCTCGCTCACCGCCACCTGTGTGTCTCCA 11190
556 -----HisGlyGlnArgHisProLeuGlnTyrSerProProIleHisGlyGlnGln 573
```

```

Qy 407 -----GlyTrpGlnIleLysGluGlnIleValSerLeu-----GluLysA 420
Db 10726 TCAGCGGTCGCGAGGAGGATATCCGGCTCCACCTCGCGTGGCCCAACGCGAGCGGG 10785
Qy 420 sPThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSerLeuSerLeu----- 437
Db 10786 TGCTGCTGGCTGTG-----CTCGAACAGCAGCGCGGGCGCTGTCTCGTCAGCATC 10836
Qy 438 -----MetGluGlnAlaAlaLeuAlaLysArgMetTyrAsnGlnIleLysArgPro- 455
Db 10837 TGGCATGACGCGCGC-----GACCGCGCGCTCTCGACGACAACTGCGCAGCGCG 10890
Qy 456 -ArgLeuSerProMetGluMetProValThrSerSerTyrSerProIleTyrAr 475
Db 10891 ACGGCATCATCTAGTCACCGGCCCCACCGCTCGGCAAGACACACCTGTACGCG 10950
Qy 475 gAspArgSerPheProSerGlnArgAspAspGlnAspGluIleSerAlaLeuValse 495
Db 10951 GCCT--GGTCACCTCAACGA----- 10969
Qy 495 rSerTyrLeuGlyProSerThrSerPheProHisArgSerArgSerProGluTyrMe 515
Db 10970 -----CCGCTCGCGCAATATCTCAGGTGAAGA-----CCGATCGAGT 11010
Qy 515 tValPro-----LeuProHisGlyGlyLeuGlyArgSerVa 527
Db 11011 ACTACCTGGAAGGCATCGGCACACCCAGGTCAACCCGCGGTGACATGACCTTCGCC 11070
Qy 527 lTyrAlaTyrGluHisLeuAlaPro-----AsnSerTyrSerProGl 541
Db 11071 GCGGCTGCGGCATCTCTCGCCAGACCCGAGCGGTGTGTGTGTGTGTGTGTGTGTGTGT 11130
Qy 541 yHisGlyHisArgLeuHisArgGlnTyrSer-----ProSerLeuVal-- 555
Db 11131 ACCAGAGACCGCGCATCGCGCTCAGGCTCGCTCAGCGCACCTGCTGTCTCTCA 11190
Qy 556 -----HisGlyGlnArgHisProLeuGlnTyrSerProProIleHisGlyGlnGln 573
Db 11191 CCTGCACACCAACAGCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 11250
Qy 573 nLeuProTyrGlyGlnArgValTyrArgHisSerProSer 587
Db 11251 CCTTCCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 11293

```

## RESULT 6

```

US-09-479-453-29
; Sequence 29, Application US/09479453
; Patent No. 6313283
; GENERAL INFORMATION:
; APPLICANT: Gerritse, Gijbert
; APPLICANT: Quax, Wilhelmus J.
; TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
; TITLE OF INVENTION: EXPRESSION LEVELS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genencor International
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/479,453
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/911,853
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: Glaister, Debra J
; REGISTRATION NUMBER: 33,888
; REFERENCE/DOCKET NUMBER: GC361-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7620
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17612 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-479-453-29

```

```

Alignment Scores:
Pred. No.: 0.0814 Length: 17612
Score: 133.00 Matches: 131
Percent Similarity: 35.95% Conservative: 98
Best Local Similarity: 20.57% Mismatches: 224
Query Match: 4.28% Indels: 184
DB: 4 Gaps: 31

```

US-09-890-475-1 (1-609) x US-09-479-453-29 (1-17612)

```

Qy 14 ThrThrAlaAsnProLeuGlnArgHisGlnSerGluGlnArgArgGluLeuPro 33
Db 9754 ACCTCGATCTCGCGTCGACCAATCGTAGCTACGCCAGCAACGCGCGGCGTG 9813
Qy 34 LysIleValGluThrGluSerThrSerMetAspIleThrIleGlyGlnSerLysGlnPro 53
Db 9814 AAGATCGCGGAGATGAGCTGCTCGCGCTTCATCGCC----- 9852
Qy 54 GlnPheLeuLysSerIleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLys 73
Db 9853 -----TCGCGCTTCGCGCACACGCGGCTGTCCTTCAAG 9885
Qy 74 ArgGlnPheAspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSerLys 93
Db 9886 CAC-----CAGGTCACAGACAAATAAGGTCATCCCTTGCTGTAACAG 9930
Qy 94 LeuGluSerAsnGlyValValLeuAlaAlaArgAsnAsn-----PheHisGln 110
Db 9931 CCGCGCGCGTGTGCGGGCTTTTGTGTGACGCGCTTACGTCCATCACACTTCTGCGCGAG 9990
Qy 111 ProMetLeuSerProArgAsnAsnValSerValGluThrThrValThrValSerGln 130
Db 9991 GTGTGCTGCGCGCT-----GCAAAATCGGCACGTGCAGTTTTCGCGC 10032
Qy 131 ProSerGlnGluIleValProGluThrSerAsnLysProGluGlyArgMetCysGlu 150
Db 10033 AAATCGGTTAACTTGGCGCTCGGCCATGCCATAAAACACACAGAACACAGCAAGATG 10092
Qy 151 -LeuMetCysSerLysGlyLeuArgLysTyrIle-----TyrAlaAsnIleSerAspG 158
Db 10093 GATCTTCTGTTCGGGAAACGATCCGCCCATGTGTCCACCGATACCCACGCCCGCTGACGG 10152
Qy 168 nAlaLysLeuMetGluGluIleProSerAlaLeuLysLeuAlaLysGluProAlaLys-P 188
Db 10153 CTCGCGAAGCCCGCTTGGCGCGCTGCGCTGCGCTTGCCTTCGCCAAACGCGCGCGTGC 10212
Qy 188 heValLeuAspCysIleGlyLysPheTyrLeuGlnGlyArgArgAlaPheThrLysGluS 208
Db 10213 TGCTGCGCAGCGCTTCGCGCAGGTCCAGTCCAGTGCAGTGCAGTGCAGTGCAGTGCAGT 10259
Qy 208 erProMetSerSerAlaArgGlnValSerLeuLeuLeuLeuLeuLeuMetP 228
Db 10260 -----GCCAGCTTGGCGCGCGTGCAGGAG-----G 10284
Qy 228 roAspArgGlyLysGlyLysValLysIleGluSerTrpIleLysAspGlnAla-----G 246
Db 10285 CCAGCGCTTCGCGCGCGCTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10344
Qy 246 luThrAlaAlaValAlaTrpArgLys-----ArgLeuMetThrGluGlyG 261

```



```

APPLICATION NUMBER: 08/699,092
FILING DATE: 16-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Glaister, Debra J
REGISTRATION NUMBER: 33,888
REFERENCE/DOCKET NUMBER: GC361-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-846-7620
TELEFAX: 650-845-6504
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 17612 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-911-853-29

Alignment Scores:
Pred. No.: 0.0814 Length: 17612
Score: 133.00 Matches: 131
Percent Similarity: 35.95% Conservative: 98
Best Local Similarity: 20.57% Mismatches: 224
Query Match: 4.28% Indels: 184
DB: 31 Gaps: 31

US-09-890-475-1 (1-609) x US-08-911-853-29 (1-17612)
QY 14 ThrThrAlaAsnProLeuLeuGlnArgHisGlnSerGluGlnArgArgGluLeuPro 33
Db 9754 ACCTCGATCTGCGCGTACCAATCGTAGTCAGCCGACAAACGGCGCAACAGCGCGGTG 9813
QY 34 LysIleValGluThrGluSerThrSerMetAspIleThrIleGlyGlnSerLysGlnPro 53
Db 9814 AAGATCGCGCAGATGAGTGCACCTCGCGCTTCATCGCC----- 9852
QY 54 GlnPheLeuLysSerIleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLys 73
Db 9853 -----TCGGCCTTCGGCGACACGGGGGTGTTCTTCAAG 9885
QY 74 ArgGlnPheAspAspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSerLys 93
Db 9886 CAC-----CAGCGTCACGAGCAATAAGGTGTCATCCCTTGCTGAACAG 9930
QY 94 LeuGluSerAsnGlyValValLeuAlaAlaArgAsnAsn-----PheHisGln 110
Db 9931 CCGCGCGCGCGTGGCGGCTTTTGTGCACGCTTACGTCCATCACACTTCGGCGCAG 9990
QY 111 ProMetLeuSerProProArgAsnAsnValSerValGluThrValThrValSerGln 130
Db 9991 GCTGTGTCGCGCT-----GCAAAATCGGCACCTGCAGTTTTTGGCGC 10032
QY 131 ProSerGlnGluLeuValProGluThrSerAsnLysProGluGlyArgMetCysGlu 150
Db 10033 AAATCCGTTAACTTGGCGCTCGGCATCGCCATGCAATAAACAACAAGCAAGATG 10092
QY 151 -LeuMetCysSerLysGlyLeuArgLysTyrlle-----TyrAlaAsnIleSerAspG 168
Db 10093 GATCTTCTGTTGGGGAACATCCGCCATGTCACCGCATCCACGCGCGCTGACGG 10152
QY 168 nAlaLysLeuMetGluGluLeuProSerAlaLeuLysLeuAlaLysGluProAlaLys-P 188
Db 10153 CTCGCCGAAGCCCGCTTCGCGCGCGCTTCGCTTCGCTTCGCAACGCCACGCGGTGC 10212
QY 188 heValLeuAspCysIleGlyLysPheTyrlleGlnGlyArgAlaPheThrLysGlu 208
Db 10213 TGCTGCGCGAGCCCTTCGGCGAGTCCAGTCGAGGTGCGCGGT----- 10259
QY 208 erProMetSerSerAlaArgGlnValSerLeuLeuIleLeuGluSerPheLeuLeuMetP 228
Db 10260 -----GCCAGCCTGCCCGCTGCAGGAG-----G 10284
QY 228 roAspArgGlyLysGlyValLysIleGluSerTrpIleLysAspGluAla-----G 246
Db 10284 -----G 246

10285 CCCAGCGCTTCGCGCGCGCGTGTGCGCTGCACCTGCTGGAGCCCGAGCCCTTCGAGC 10344
QY 246 luThrAlaAlaValAlaIleArgLys-----ArgLeuMetThrGluGlyG 261
Db 10345 AGGAGCTGCGCTTCGCGCTTACCGCGCGCGACTCTCTCGAGGTGCGGCAGATGCGCGAGGSC- 10403
QY 261 lYleuAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuValAlaCysPheG 281
Db 10404 -----ATGGGTGCGCAACTTGACCTAGCCAGCTTGGCGCACTCACTCCC----- 10448
QY 281 lYValProSerAsnPheArgSerThrAspLeuLeuAspLeuIleArgMetSerGlySerA 301
Db 10449 -----GAAATCCGCGGACCTTCTGGAG----- 10469
QY 301 snGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetValSerGlyIleV 321
Db 10470 -----CAGGAAGATGACGCGCGCGATCATCCGCTGATCAACGCCATCC 10512
QY 321 alGluSerSerIleLysArgGly-----MetHisIleGluAlaLeuGlu-----M 336
Db 10513 TCAGCGAGCGCATCAAGCGCGCGCTCCGACATCCACCTGGAAACCTTCGAGAAACGCC 10572
QY 336 etValTyThrPheGlyMetGluAspLysPheSerAlaAlaLeuValIleThrSerPheL 356
Db 10573 TGGTGTGCGCTTTCGCGTCGACGCG-----ATCCTCGCGCAAGTGA 10614
QY 356 euLysMetSerLysGluSerPheGluArgAlaLysArgLysAlaGlnSerProLeuAlaP 376
Db 10615 TCGAAGCGCGCGCGAG-----CTGCGC- 10637
QY 376 heLysGluAlaAlaThrLysGlnLeuAlaValLeuSerSerValMetGlnCysMetGluT 396
Db 10638 -----GGCTGCTGCTTCGCGGTCAAGGTCAATGCGG- 10670
QY 396 hrHisLysLeuAspProLa-----LysGluLeuPro----- 406
Db 10671 -----CGCTGTGACATCGCGAGAGCGGTACGCGAGCGCGCTATTTCGCTCAAGG 10725
QY 407 -----GlyTrpGlnIleLysGluGlnIleValSerLeu-----GluLysA 420
Db 10726 TCGCGGTGCGCGAGTGTATTCGCGTCTCCACCTCGCTCGGCCAACGCGGAGCGGG 10785
QY 420 spThrLeuGlnLeuAspLysGluMetGluLysAlaAlaArgSerLeuSerLeu----- 437
Db 10786 TGTGTGCTGCTG-----CTCGACAAGAGCGCGCGCGCTGTGCTCACGCAATC 10836
QY 438 -----MetGluGluAlaAlaLeuAlaLysArgMetTyrAsnGlnGlnIleLysArgPro- 455
Db 10837 TGGGCATGAGCGAGCGC-----GACCGCGCGCTGCTCGAGCAACCTGCGCAAGCGCG 10890
QY 456 -ArgLeuSerProMetGluMetProValThrSerSerSerTyrSerProIleTyAr 475
Db 10891 ACGGCATCATCTAGTACCGCGCGCGCGCTCGGCGCAAGACCAACCCCTGTACGCGC 10950
QY 475 gasArgSerPheProSerGlnArgAspAspGlnAspGluIleSerAlaLeuValse 495
Db 10951 GCCT---GGTCAACCTCAACGA----- 10969
QY 495 rSerTyLeuGlyProSerThrSerPheProHisArgSerArgArgSerProGluTyMe 515
Db 10970 -----CCGCTCGCGCATATCTTCACGGTGAAGA-----CCGATCGAGT 11010
QY 515 tValPro-----LeuProHisGlyGlyLeuGlyArgSerVa 527
Db 11011 ACTACCTGAAGGATCGGCCAGACCCAGGTCAACCGCGCGGTGACATGACCTTCGCC 11070
QY 527 lTyAlaTyGluHisLeuAlaPro-----AsnSerTySerProGl 541
Db 11071 GCGGCTCGCGCGCATCTTCGCCCGAGGACCGGACGCTGTGATGTCGCGGAGATCGCG 11130
QY 541 yHisGlyHisArgLeuHisArgGlnTySer-----ProSerLeuVal-- 555
Db 11131 ACCAGGAGACCGCGGACATCGCGGTGCGAGGCGCTCGCTCACGCGGACCTGTGCTCCA 11190

```

2y 189 ValLeuAspCysIleGlyLysPheTyrLeuGlnGlyArg-----AlaPheThr 205  
Db 4330 GTATTAAGAGCTGTGGGAGAACTGGAAGAGGTTCAGCACCTGGGCTCTTCTCCCGT 4389  
2y 206 LysGluSerProMetSerAlaArgGlnValSerLeuLeuLeuLeuLeuLeuLeu 225  
Db 4390 AAGAACGAGATCCACAGAAACAGACAGAGATGCTCTGTGTAGCCACAGCCTTTCTTA 4449  
2y 226 -----LeuMetProAspArgGlyLysValLysValLysLysLysLysLys 240  
Db 4450 TCACTGGTGTAAATGCTGAAAGGACAGATGAGCAGCGGAGGAGCTCTCGGCTCC 4509  
2y 241 IleLysAspGluAlaGluThrAlaAlaValAlaValArg----- 253  
Db 4510 CTCACAGCGGAGGTCACGAGATGTTGATTAAATTGGCGAGAAACCAGTACTAGATGAT 4569  
2y 254 -----LysArgLeuMetThrGlu-----GlyGly 261  
Db 4570 TGCAAAACCAAGCAGCTAAATGATGAGGCTCAAACTGGCGGTCAACCTGGTGGGGGC 4629  
2y 262 Leu-----AlaAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeu 275  
Db 4630 ATGTTTGACACTGTGACGCTAGTAGTACCCAGCAGCTACAGAGTGGCGCCAGCTTCCTT 4689  
2y 276 LeuValAlaCysPheGlyVal-----ProSerAsnPheArgSerThrAspLeu 291  
Db 4690 GAGATCATCATCAGCGGACCTGTGGACATGAGTCAAGCTAAACATGAGCTCTTCACTACTGTG 4749  
2y 292 LeuAspLeuIleArgMetSerGlySerAsnGluLeuAlaGlyAlaLeuLysArgSerGln 311  
Db 4750 TTGACATCTGAGCGTCTTATCAACGGAACGTTGGCTCCAGACATG----- 4797  
2y 312 PheLeuValProMetValSerGlyLeuValGluSerSerIleLysArgGlyMetHisIle 331  
Db 4798 -----TCCAGTATCTCGCAAGC----- 4815  
2y 332 GluAlaLeuGluMetValTyrPheGlyMetGluAspLysPheSerAlaAlaLeuVal 351  
Db 4816 -----AGCATGGAGGAAACCAAGGCTGCATATATGAAC 4848  
2y 352 LeuThrSerPheLeu-----LysMetSerLysGluSerPheGluArgAla 366  
Db 4849 CTGGTGAAGAGCTTCAGAGGACTTGGGGAGCGCCCAATCAGACAGTCTGGAGAAGTT 4908  
2y 367 LysArgLysAlaGlnSerProLeuAlaPheLysGluAlaAlaThr----- 381  
Db 4909 CACCAACTGTTGCCACTACCCAGCAGAACCGAGATGTCATAACCTGTGAGCCACAGGCG 4968  
2y 392 -----LysGluLeuAlaValLeuSerSerVal----- 390  
Db 4969 TCCCTTATGACCAAGGGCAACAAAGATTGCTGGCTTCGATTCCATCTTCAAGAAGGAG 5028  
2y 391 ---MetGlnCysMetGluThrHisLysLeuAspProAlaLysGluLeuProGlyTTPGln 409  
Db 5029 GGTCTACAAGTTTACCAACAAAGATCTCTCC-----TGGAG 5070  
2y 410 IleLysGluGlnIleValSerLeuLys-----AspThrLeu 422  
Db 5071 CTTTGTAGGCGCTGAAGCCATCAACAGCACCCTGTCATGGGCGCTGGTTGSCACAGTC 5130  
2y 423 GlnLeuAspLysGluMet-----GluGluLysAlaArgSerLeuSerLeuMetGlu 439  
Db 5131 CGAGTGGACCGCAGAGTGGCAGGAGGAGGAGCAGCGGCTGTG----- 5178  
2y 440 GluAlaAlaLeuAlaLysArgMetTyrAsnGlnIleLys---ArgProArg----- 456  
Db 5179 -----CTCATCATACCACCTGAGCGCTGACCCAGAGCCTAT 5217  
2y 457 ---LeuSerProMetGluMetProValThrSerSerSerTyrSerProIleTyr--- 474  
Db 5218 TACCTGGAACCACTACCTCTGCCCCAGAGATGAGGAGCCAGCGCTGCCCTACTA 5277  
2y 474 ----- 474

Db 5278 GAGCCTGAGAAAAAGGCTCCTGAGCCCCCAAGACTGACAAACACGAGGGGCTGCTCCTCG- 5337  
Qy 475 -----ArgAspArgSerPheProSerGln 482  
Db 5338 AGCACTGAGGAGCGCAAAAGAGTCTCAAGGGCAAAAACGACGACGACCCACC 5397  
Qy 483 ArgAspAspGlnAspGlnLeuSerAlaLeuValSerTyrLeuGlyProSerThr 502  
Db 5398 AAGAACGAGACTATGCG-----ATGGGGCCAGGTCGG 5430  
Qy 503 SerPheProHisArgSerArgSerProGluTyrMet----- 515  
Db 5431 AGTGGCCCTATGTTGACAGTGCCTCCAGACCTTCTACACCATGCAAACTCCTGTTCT 5490  
Qy 516 -----ValProLeu 518  
Db 5491 ATATCCACCTTAGCTACAGGCAAGCTCCATGGGCTGTATACCCAAACACGACCTA 5550  
Qy 519 ProHisGlyGly----- 522  
Db 5551 CCTGCTGGTGGCCCTCGTGTGGATCCATACCGCCGCGTCCGATTACCAATGCAAAAGCTG 5610  
Qy 523 -----LeuGlyArgSerValTyrAlaTyrGluHisLeu 533  
Db 5611 CCAACTCGACCAACTTATCCCGGTGTGCTGCCTCAACAATGCTACTGTCTATGTCGTCCTA 5670  
Qy 534 AlaProAsnSerTyrSerProGlyHisGlyHisArgLeuHisArgGlnTyrSerProSer 553  
Db 5671 GAACCTCTTCTTATAAGACATCT-----GTATACCGGACGAGCAGCAACCCACA 5718  
Qy 554 LeuValHisGlyGln-----ArgHisProLeuGlnTyrSerProIleHisGlyGln 571  
Db 5719 GTGCCCCAGGACAGCGCTTCGCCAACAGCTCCAGCAGTCCAGGGATGTTGGACAG 5778  
Qy 572 -----GlnGlnLeuPro-----TyrGlyIleGlnArgValTyrArgHis 584  
Db 5779 TCATCTGTCATCAGATGACCCCTAGTCTTCTATGTTTGCAGACTTCCAGCTCTCT 5838  
Qy 585 SerProSerGluGluArgTyr-----LeuGlyLeuSerAsnGlnArg 598  
Db 5839 TCCTCTCTCTCCAGGGCTATACCTATGTTTCTCATGTGGATTGCGACACACACA 5898  
Qy 599 SerPro 600  
Db 5899 GGCCT 5904  
RESULT 4  
US-08-911-853-29  
; Sequence 29, Application US/08911853  
; Patent No. 6048710  
; GENERAL INFORMATION:  
; APPLICANT: Gerritse, Gijbert  
; APPLICANT: Quax, Wilhelmus J.  
; TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED  
; TITLE OF INVENTION: EXPRESSION LEVELS  
; NUMBER OF SEQUENCES: 37  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genencor International  
; STREET: 925 Page Mill Road  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304-1013  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/911,853  
; FILING DATE:  
; PRIOR APPLICATION DATA:

Qy 318 SerGlyIleValGluSerSerIleLysArgGlyMetHisIleGluAlaLeuGluMetVal 337  
Db 2209 -----GAACGGAACTAAGACGG-----AAACATGTC----- 2235  
Qy 338 TyrThrPheGlyMetGluAspIlePheSerAlaLeuValLeuThrSerPheLeuLys 357  
Db 2236 -----ATGAAGTTGCACGACGCTAGAGTTGAAGTCTAAAGAACTCA 2283  
Qy 358 MetSerLysGluSerPheGluArgAlaLysArgLysAlaGlnSerProLeuAla 375  
Db 2284 ATAAAAAGCAGTTTCAGGATACCTGCAAAATTCAAACACAGACAGTACAAAGCAATTAAGG 2343  
Qy 376 -----PheLys 377  
Db 2344 AATCACTACTGAGACTACCAAAAGAGTGAGCACAAGAGCTGTCTGAAAAGACTCAAG 2403  
Qy 378 GluAlaAlaThrLysGlnLeuAlaValLeuSer-----SerValMetGln 392  
Db 2404 GAGGAACAGACTCGGAAGTTAGCATCTTGGCTGAGCAGTATGATCATAGCATTAATGAA 2463  
Qy 393 CysMetGluThrHis-----LysLeuAspProAlaLysGlu----- 404  
Db 2464 ATGCTCTCCACACAGCTCTGGCTTGGATGAGCAGCAGAGCAGAGATGCCAGGTTTG 2523  
Qy 405 -----LeuProGlyTyrGlnIleLys----- 411  
Db 2524 AAGATGACGCTACAGCAGGAAGCTGGAGCTGTGTAATGCATACAGACAAATCAAGATG 2583  
Qy 412 -----GluGlnIleValSerLeu----- 417  
Db 2584 CAGGCTGAGGCCCAACATGATCAGAGCTTCGAGAGCTGGAACAAGGCTCTCCCTTCGG 2643  
Qy 418 -----GluLysAspThrLeuGlnLeuAspLysGluMetGlu 429  
Db 2644 AGAGCACTCTTAGACAGAGATTGAAGAGAGATGTTGGCTTCAGAAATCAAGCACA 2703  
Qy 430 GluLysAlaArgSerLeuSerLeuMetGluGluAlaLeuAlaLysArgMetTyrAsn 449  
Db 2704 GAACGATACGT-----AGCCTGCTCGAGCGCCAGGCGAGAAATTAAGACTTTTGC 2757  
Qy 450 GlnGlnIleLysArgProArgLeuSerProMetGluMetProValThrSerSerSer 469  
Db 2758 TCTGAAGCATGAGATTAGATTGTTTGAATACATGCTCTT-----TCTAAT 2802  
Qy 470 TyrSerProIleTyrArgAspArgSerPheProSerGlnArgAspAspGlnAspGlu 489  
Db 2803 CTCTCCCTGAGCATTCAGCCACAGCTACCCAGGA----- 2838  
Qy 490 IleSerAlaLeuValSerSerTyrLeuGlyProSerThrSerPheProHisArgSerArg 509  
Db 2839 -----GCTTCTAGTGTCTCACAAT----- 2859  
Qy 510 ArgSerProGluTyrMetValProLeuProHisGlyLeuGlyArgSerValTyrAla 529  
Db 2860 -----CCTACTGGGGGTTTCAGGA----- 2877  
Qy 530 TyrGluHisLeuAlaProAsnSerTyrSerProGlyHisGlyHisArgLeuHisArgGln 549  
Db 2878 -----CCTCACTGGGGTCACTCCCAIG----- 2898  
Qy 550 TyrSerProSerLeuValHisGlyGlnArgHisProLeuGlnTyrSerPro----- 566  
Db 2899 -----GTTGSCACACCAAGCTTGGGTCATCCGATGCAAGCGGACCCCAACCATGG 2952  
Qy 567 -----ProIleHisGlyGlnGlnLeuProTyrGlyIleGlnArgValTyrArgHis 584  
Db 2953 GGTCAACCCCTCAGGGCCCAATGCAAGGGGTACCTCGAGGTAGCATAGATGATCGCAT 3012  
Qy 585 SerProSerGluGluArgTyrLeu-----GlyLeuSerAsnGlnArgSerProArg 601  
Db 3013 AGCCCCCAGGCTCTGAGGCGGACAGCTTCTGGGGGACGAGCGGACAGGCGCATGAGCAGA 3072  
Qy 602 SerAsnSer 604

Db 3073 AGCAGAGT 3081  
RESULT 3  
US-09-491-356C-7  
; Sequence 7, Application US/09491356C  
; Patent No. 6566061  
; GENERAL INFORMATION:  
; APPLICANT: Philibert, Robert A.  
; APPLICANT: Gihns, Edward I.  
; TITLE OF INVENTION: IDENTIFICATION OF POLYMORPHISMS IN THE PCTG4 REGION OF XQ13  
; FILE REFERENCE: 9465.6US11  
; CURRENT APPLICATION NUMBER: US/09/491.356C  
; CURRENT FILING DATE: 2000-01-26  
; PRIOR APPLICATION NUMBER: PCT/US99/09365  
; PRIOR FILING DATE: 1999-04-29  
; PRIOR APPLICATION NUMBER: 60/083,465  
; PRIOR FILING DATE: 1998-04-29  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 7  
; LENGTH: 6558  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-09-491-356C-7  
Alignment Scores:  
Pred. No.: 0.0055 Length: 6558  
Score: 137.50 Matches: 137  
Percent Similarity: 33.45% Conservative: 118  
Best Local Similarity: 17.98% Mismatches: 226  
Query Match: 4.43% Indels: 281  
DB: 36  
US-09-890-475-1 (1-609) x US-09-491-356C-7 (1-6558)  
Qy 17 AsnProLeuLeuGlnArgHisGlnSerGluGlnArgArgGluLeuProLysIleVal 36  
Db 3928 GACCCAGTGTGAGTAGTGCCAGGCCCGCCGCGC-----CTCATGCAACTATC 3975  
Qy 37 GluThrGluSerThrSerMetAspIleThrIleGlyGlnSerLysGlnProGln----- 54  
Db 3976 TGCTACCCACATCGACTGCTGCAATCAGGATGAGAAACCCAGCGCGCGCAT 4035  
Qy 55 -----PheLeuLysSerIleAspGluLeuAlaPheSerValAlaValGluThrPhe 72  
Db 4036 AAACGTATTCTCAAGATTTAGCAATGG-----ACCATG 4071  
Qy 73 LysArgGlnPheAspAspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSer 92  
Db 4072 CGCCAGTCTCTTTGGAGCTACAGTTGATGATCAAGCAGACCCCAACACT----- 4122  
Qy 93 LysLeuGluSerAsnGlyValValLeuAlaAlaPheSerValAlaValGluThrProMet 112  
Db 4123 -----GAGATGAATCTCTCTTG----- 4140  
Qy 113 LeuSerProArgAsnAsnValSerValGluThrValThrValSerGlnProSer 132  
Db 4141 -----GAGACATGCCC-----AAGGCCACAAATCGAGTTTCCAAAGTCA 4182  
Qy 133 GlnGluIleValProGluThr-----SerAsnLysProGluGlyGlyArgMet 148  
Db 4183 GCAGACAGCGGCTCATCTTCTGGAAGTACAGCAACATGCGCCAGCAGCAAGACC 4242  
Qy 149 CysGluLeuMetCysSerLysGlyLeuArgLysTyrIleTyrAlaAsnIleSerAspGln 168  
Db 4243 AAACCTGTGCTCAGTCT-----CTAGAGCGA-----TCTGGTGTATGTTGGTG 4287  
Qy 169 AlaLysLeuMetGluGluLeuProSerAlaLeuLysLeuAlaLysGluProAlaLysPhe 188  
Db 4288 GCTCTCTCATTCGCAAACTGCCCACTTCAGTCCAG-----GGCCAT 4329

```

2878 -----CCTCACTGGGGTCAATCCCATG----- 2898
550 TyrSerProSerLeuValHisGlyGlnArgHisProLeuGlnTyrSerPro----- 566
2899 -----GGTGGCACACACCAAGCTTGGGGTCAATCCCATGCAAGCGGACCCCAACCATGG 2952
567 -----ProTleHisGlyGlnGlnLeuProTyrGlyGlnArgValTyrArgHis 584
2953 GGTCAACCCCTCAGGCGCAATGCAAGGGTACCTCGAGGTAGCAGTATAGAGTCCGCAAT 3012
585 SerProSerGluGluArgTyrLeu-----GlyLeuSerAsnGlnArgSerProArg 601
3013 AGCCCCCAGGCTCTGAGCGGACAGCTTCTGGGGGACGCGACGCAAGCGGCATGACGACA 3072
602 SerAsnSer 604
3073 AGCAGCAGT 3081

RESULT 2
JS-09-723-458-1
: Sequence 1, Application US/09723458
: Patent No. 6586242
: GENERAL INFORMATION:
: APPLICANT: Cobb, Melanie
: Hutchinson, Michele
: Chen, Zhu
: Berman, Kevin
: TITLE OF INVENTION: TAO PROTEIN KINASES AND METHODS OF USE
: THEREFOR
: NUMBER OF SEQUENCES: 26
: CORRESPONDENCE ADDRESS:
: ADDRESS: SEED and BERRY LLP
: STREET: 6300 Columbia Center, 701 Fifth Avenue
: CITY: Seattle
: STATE: Washington
: COUNTRY: USA
: ZIP: 98104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/723,458
: FILING DATE: 27-No. 6586242-2000
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/09/060,410
: FILING DATE: <Unknown>
: ATTORNEY/AGENT INFORMATION:
: NAME: Maki, David J.
: REGISTRATION NUMBER: 31,392
: REFERENCE/DOCKET NUMBER: 860098.421
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 622-4900
: TELEFAX: (206) 692-6031
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3312 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 121..3123
: SEQUENCE DESCRIPTION: SEQ ID NO: 1:
JS-09-723-458-1

Alignment Scores:
Pred. No.: 7,64e-05 Length: 3312
Score: 151.00 Matches: 150
Percent Similarity: 34.75% Conservative: 95
Best Local Similarity: 21.28% Mismatches: 237

```

```

Query Match: 4.86% Indels: 223
DB: 4 Gaps: 32
US-09-890-475-1 (1-609) x US-09-723-458-1 (1-3312)

QY 3 AsnTyr-----ProProThrValAlaAlaGlnProThrThrAlaAsn 17
1333 AATTACCAAGAGAAGAGATCTTAGAACAGAGATCGCTCCACAGTCTCCA----- 1386
18 ProLeuLeuGlnArgHisGlnSerGluGlnArgArgGluLeuProLysIleValGlu 37
1387 CCTCAAGTGCTCGTCACAAATCACATTATCGTAATAGAGAACACTTTTGCACACTATAGA 1446
38 ThrGluSerThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLys 57
1447 ACAGCATCACTG-----GTTACAAGACAGATGCAAGAACATGAG 1485
58 SerIleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAsp 77
1486 CAGGACTCTGAACCTTAGAGAACAGATGTCTGTTATTAAGCGGATGAGCGACAGCAT-- 1542
78 AspLeuGlnLysIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsn 97
1543 -----CAGAAGCAGCTGATGACTCTGGAATAATAAATCTGAAGGCAGAAATGACGAACAT 1596
98 GlyVal-----ValLeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeu 113
1597 CGGCTCAGATTAGACAAAGATCTTGAACCTCAGCGCAACAATTCGCTGCAGAAATG-- 1653
114 SerProProArgAsnAsnValSerValGluThrThrValThrValSerGlnProSerGln 133
1654 -----GAGAACTTATTAAAGAAACACCAAGCTTCTTATG 1686
134 GluIleValProGluThrSerAsnLysProGluGlyGlyArgMetCysGluLeuMetCys 153
1687 GAAAAGAGGCTAAAGTATGCGCCACAGAGGAG----- 1719
154 SerLysGlyLeuArgLysTyrIleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGlu 173
1720 -----AAAAAATTCACAAACACAT-----CAGGCTCACAGAGAGAA 1758
174 GluIleProSerAlaLeuLys-LeuAlaLysGluProAlaLysPheValLeuAspCysIle 193
1759 GAATGTAATAGCTTTTGGAGTCTCAAAAAGAGAAATATAAATCTCGAAAAGAGAGAGCTT 1818
193 eGlyLysPheTyrLeuGlnGlyArgAlaPheThrLysGluSer-----ProMetSe 211
1819 AAGGAGGAGCTGAATGAAACCCAGACACACCTTAAAAAAGAAAGCAGGATGGCTTTCA 1878
211 rSerAlaArgGlnValSerLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 231
1879 AAGCAGAAGAGAGAAATATTCACATTTTCAGGCGAGAGAGAGAGAGCTAATCTTCTCGAGCT 1938
231 YLysGlyLysVal-----LysIleGluSerTrpIleLysAspGluAlaGluTh 247
1939 CAAGGAGATATCTAGAGCTAGATGCTCGCTTCA-----AAAGAAGATGTTACTT 1992
247 rAlaAlaValAlaTrpArgLysArgLeuMetThrGlu-GlyGlyLeuAlaAlaGluL 267
1993 GGTGGCATAACTTGAACAGGACCTTG-----TCAGGAGGAGGT-----AAACAAA 2040
267 yMetAspAlaArgGlyLeuLeuLeu-----LeuValAlaLac 279
2041 AGCGAGACTCAGAGAGGACTTAGAACATGCAATGTTACTCGGACAGCATGAATCCATGCA 2100
279 ySPheGlyVal---ProSerAsnPheArgSerThrAspLeuLeu-AspLeuIleArgMet 297
2101 GAATGGAGTTTCGCCACCTCAACACTATTTCAGAGATGCGCTGTGAGTTGATCAGACTG 2160
298 SerGlySerAsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetVal 317
2161 CAACATCAAACTGAGCTTACTTAACAGCTGGGAATACAAATAGAGAGAGG----- 2208

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 3312 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 121..3123
US-09-060-410-1

Alignment Scores:
Pred. No.: 7,64e-05 Length: 3312
Score: 151.00 Matches: 150
Percent Similarity: 34.75% Conservative: 95
Best Local Similarity: 21.28% Mismatches: 237
Query Match: 4.86% Indels: 223
DB: 32 Gaps: 32

US-09-890-475-1 (1-609) x US-09-060-410-1 (1-3312)
QY 3 AsnTyr-----ProProThrValAlaAlaGlnProThrThrThrAlaAsn 17
DB 1333 AATTACCAAGAGAGAGATCCTAGACAAGAGCATCAGCTCCACAGTCTCCA-----1386
QY 18 ProLeuLeuGlnArgHisGlnSerGluGlnArgArgGluLeuProLysIleValGlu 37
DB 1387 CCTCAAGTGTCTCTGCACAAATCACATTATCGTAATAGAGACACTTTGCCAACTATACGA 1446
QY 38 ThrGluSerThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLys 57
DB 1447 ACAGCATCACTG-----GTTACAAAGACAGATGAGCGAATGAG 1485
QY 58 SerIleAspGluLeuAlaPheSerValAlaValGluThrPheLysArgGlnPheAsp 77
DB 1486 CAGGACTCTGAATTTAGAACACAGATGCTGTTATAGCGGATGAGCGACAGCAT---1542
QY 78 AspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsn 97
DB 1543 -----CAGAAAGCAGCTGATGACTCTCGAAATATAAATCTGAAGCGAATAATGACGAACAT 1596
QY 98 GlyVal-----ValIleuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeu 113
DB 1597 CGGCTCAGATTAGAACAAAGATCTTGAATCTAGCCGCAACATTTTCGTCGCAAAATG---1653
QY 114 SerProProArgAsnValSerValGluThrThrValThrValSerGlnProSerGln 133
DB 1654 -----GAGAAACTTATTAAAGAAACACCAAGCTTCTATG 1686
QY 134 GluIleValProGluThrSerAsnLysProGluGlyGlyArgMetCysGluLeuMetCys 153
DB 1687 GAAAAAGAGCTAAAGTGATGGCCACAGAGGAG-----1719
QY 154 SerLysGlyLeuArgLysTyrIleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGlu 173
DB 1720 --AAAAAATCCAACACACATTT-----CAGGCTCAACAGAGAAGAA 1758
QY 174 GluIleProSerAlaLeuLys-LeuAlaLysGluProAlaLysPheValLeuAspCysI1 193
DB 1759 GAACTGAATAGCTTTTGGAGTCTCAAAAAGAGAGATATAAACTTCGAAAAAGACAGCTT 1818
QY 193 eGlyLysPheTyrLeuGlnGlyArgAlaPheThrLysGluSer-----ProMetSe 211
DB 1819 AAGGAGAGCTGTAAGAAACACAGACACACCTAAAAAAGAAAGACGGAATGGCTTTCA 1878
QY 211 rSerAlaArgGlnValSerLeuLeuIleLeuGluSerPheLeuLeuMetProAspArgG1 231
DB 1879 AAGCAGAGAGAGAGATATCAACATTTTCAGCGCAGAGAGAGAGCTAAATCTTCTCGACGT 1938
QY 231 YLysGlyLysVal-----LysIleGluSerTrpIleLysAspGluAlaGluTh 247
DB 1939 CAAGGACGATATCTAGAGCTAGATGCTCGCTTCA-----AAAGAGAGATGTTACTT 1992
QY 247 rAlaAlaValAlaTrpArgLysArgLeuMetThrGlu-GlyGlyLeuAlaAlaGluL 267
;
1993 GGTGGCATAAAGTGGAAACAGGACCTTG-----TCAGGAGGAGTT-----AAACAAA 2040
QY 267 yMetAspAlaArgGlyLeuLeuLeu-----LeuValAlaC 279
DB 2041 AGGAGACTCAGAAGGACTTAAACATGCAATGTTACTGCGACAGCATGATCATGCAAA 2100
QY 279 yPheGlyVal---ProSerAsnPheArgSerThrAspLeuLeu-AspLeuIleArgMet 297
DB 2101 GAACTGGAGTTTGGCCACCTCAACACTATTCAAGAGATGCGCTGTGAGTTGATCAGACTG 2160
QY 298 SerGlySerAsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetVal 317
DB 2161 CAACATCAAACTGAGCTTACTTAACAGCTGGAATACAAATAGAGAGG-----2208
QY 318 SerGlyIleValGluSerSerIleLysArgGlyMethHisIleGluAlaLeuGluMetVal 337
DB 2209 -----GAACGGGAACTAAGACGG---AAACATGTC-----2235
QY 338 TyrThrPheGlyMetGluAspLysPheSerAlaAlaLeuValLeuThrSerPheLeuLys 357
DB 2236 -----ATGGAAGTTTCGACAGCAGCCTAAGAGTTTGAAGTCTAAAGACTCCAA 2283
QY 358 MetSerLysGluSerPheGluArgAlaLysArgLysAlaGlnSerProLeuAla-----375
DB 2284 ATAAAAAGCAGTTTCAGGATACCTGCAAAATTCAAACACAGACAGTACAAAGCATTAAGG 2343
QY 376 -----PheLys 377
DB 2344 ATACCTACTGGAGACTACCAAGAGTGAGCACAAAGCTGTTCTGAAAAGACTCAAG 2403
QY 378 GluAlaAlaThrLysGlnLeuAlaValLeuSer-----SerValMetGln 392
DB 2404 GAGGAACAGACTCGGAAGTTAGCCATCTTGCTGAGCAGTATGATCATGATTAATGAA 2463
QY 393 CysMetGluThrHis-----LysLeuAspProAlaLysGlu-----404
DB 2464 ATGCTCTCCACACAAAGCTCTGCTGTTGATGAAGCACAGGAGCAGAAATCCAGGTTTG 2523
QY 405 -----LeuProGlyTrpGlnIleLys-----411
DB 2524 AAGATCCAGCTACACAGGAACTGGAGCTGTTGAATGATATATCAGAGCAAAATCAAGATG 2583
QY 412 -----GluGlnIleValSerLeu---417
DB 2584 CAGGCTGAGGCCCAACATGATCGAGAGCTTCGAGAGCTGGAAACAAAGGGTCTCCCTTCGG 2643
QY 418 -----GluLysAspThrLeuGlnLeuAspLysGluMetGlu 429
DB 2644 AGAGCACTCTTAGAACAGAAAGATTGAAGAGAGATGTTGGCTTTCAGAAATGAACGCACA 2703
QY 430 GluLysAlaArgSerLeuSerLeuMetGluGluAlaLeuAlaLysArgMetTyrAsn 449
DB 2704 GAACGAATACGT-----AGCTGCTCGAGCGCCAGAGAAATGAAGCTTTTGAC 2757
QY 450 GlnGlnIleLysArgProArgLeuSerProMetGluMetProProValThrSerSerSer 469
DB 2758 TCTGAAGCATGAGATTAGTTAGTTAGTAACTGTCCTT-----TCTAAT 2802
QY 470 TyrSerProIleTyrArgAspArgSerPheProSerGlnArgAspAspGlnAspGlu 489
DB 2803 CTCTCCCTGAGGCATTTCAGCACAGCTTACCAGGA-----2838
QY 490 IleSerAlaLeuValSerSerTyrLeuGlyProSerThrSerPheProHisArgSerArg 509
DB 2839 -----GCTTCTAGCTGGTCTCAAT-----2859
QY 510 ArgSerProGluTyrMetValProLeuProHisGlyGlyLeuGlyArgSerValTyrAla 529
DB 2860 -----CCTACTGGGGGTTTCAGGA-----2877
QY 530 TyrGluHisLeuAlaProAsnSerTyrSerProGlyHisGlyHisArgLeuHisArgGln 549

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

DM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 25, 2003, 10:24:14 ; Search time 104 Seconds

(without alignments)  
2584.638 Million cell updates/sec

Title: US-09-890-475-1

Perfect score: 3104

Sequence: 1 MSNPPVAAQPTTANPLL.....RYLGSLNQRSPRNSGLDPK 509

Scoring table:

	BLOSUM62
Xgapop 10.0, Xgapext 0.5	
Ygapop 10.0, Ygapext 0.5	
Fgapop 6.0, Fgapext 7.0	
Dgapop 6.0, Delcxt 7.0	

Searched: 559978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

```
-MODEL=frame+p2n model -DEV=xlh
-Q=/cgn2_1/USPTO/spool/US09890475/runat 14082003 090351 19610/app query.fasta_1.775
-DB=Issued Patents NA -QMWI=fastcap -SUFFIX=rni -MINMATCH=0.1 -LOOFCI=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=Blosum62 -TRANS=human0.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09890475 @CGN 1 1 56 @runat 14082003 090351 19610 -NCPU=6 -ICPU=3
-NO WMAP -LARGEROVER -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
```

Database : Issued Patents NA:\*

```
1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	151	4.9	3312	3	US-09-060-410-1
2	151	4.9	3312	4	US-09-723-458-1
3	137.5	4.4	6558	4	US-08-491-356C-7
4	133	4.3	17612	3	US-08-911-853-29
5	133	4.3	17612	3	US-09-479-409-29
6	133	4.3	17612	4	US-08-479-453-29
7	131	4.2	9775	3	US-08-977-171-1
8	130.5	4.2	7393	4	US-09-620-312D-372
9	129	4.2	3324	4	US-09-620-312D-1020
10	127.5	4.1	3605	3	US-09-098-901-1
11	127	4.1	6306	1	US-08-466-390-3
12	127	4.1	6306	1	US-08-470-950-3

13	127	4.1	6306	1	US-08-467-781-3	Sequence 3, Appli
14	127	4.1	6306	1	US-08-195-487-3	Sequence 3, Appli
15	127	4.1	6306	5	PCT-US93-06160-3	Sequence 3, Appli
16	127	4.1	6306	5	PCT-US93-06160-3	Sequence 3, Appli
17	126.5	4.1	9934	3	US-08-977-171-2	Sequence 2, Appli
18	126.5	3.9	7100	4	US-09-308-375-1	Sequence 1, Appli
19	120	3.9	11220	3	US-09-105-537-32	Sequence 32, Appli
20	120	3.9	36778	3	US-09-105-537-5	Sequence 5, Appli
21	119.5	3.8	5053	2	US-08-685-576-2	Sequence 2, Appli
22	118	3.8	38506	3	US-09-320-878-19	Sequence 19, Appli
23	118	3.8	38506	4	US-09-141-908-1	Sequence 19, Appli
24	118	3.8	38506	4	US-09-657-440-19	Sequence 19, Appli
25	117.5	3.8	2487	1	US-08-261-304-1	Sequence 1, Appli
26	116.5	3.8	4403765	3	US-09-103-840A-2	Sequence 2, Appli
27	116.5	3.8	4411529	3	US-09-103-840A-1	Sequence 1, Appli
28	116	3.7	1512	3	US-08-911-853-6	Sequence 6, Appli
29	116	3.7	1512	3	US-09-479-409-6	Sequence 6, Appli
30	116	3.7	1512	4	US-09-479-453-6	Sequence 6, Appli
31	116	3.7	2680	2	US-08-533-306A-5	Sequence 5, Appli
32	116	3.7	2680	2	US-08-742-923A-5	Sequence 5, Appli
33	116	3.7	2887	2	US-08-533-306A-3	Sequence 3, Appli
34	116	3.7	2887	2	US-08-742-923A-3	Sequence 3, Appli
35	116	3.7	7453	4	US-09-620-312D-248	Sequence 248, App
36	116	3.7	7501	4	US-09-620-312D-249	Sequence 249, App
37	116	3.7	10136	1	US-08-353-700-2	Sequence 2, Appli
38	116	3.7	10136	5	PCT-US95-16216-2	Sequence 2, Appli
39	115.5	3.7	2859	4	US-09-252-991A-944	Sequence 944, App
40	115	3.7	3378	4	US-09-513-783A-151	Sequence 151, App
41	115	3.7	4833	4	US-09-513-783A-21	Sequence 21, Appli
42	115	3.7	8789	1	US-08-328-254-5	Sequence 5, Appli
43	114.5	3.7	1836	4	US-09-252-991A-15831	Sequence 15831, A
44	114.5	3.7	2055	4	US-09-252-991A-15771	Sequence 15771, A
45	114.5	3.7	4080	2	US-08-446-345-35	Sequence 35, Appli

#### ALIGNMENTS

```
RESULT 1
US-09-060-410-1
; Sequence 1, Application US/09060410
; Patent No. 6165461
; GENERAL INFORMATION:
; APPLICANT: Cobb, Melanie
; APPLICANT: Hutchinson, Michele
; APPLICANT: Chen, Zhu
; APPLICANT: Berman, Kevin
; TITLE OF INVENTION: TAO PROTEIN KINASES AND METHODS OF USE
; TITLE OF INVENTION: THEREFOR
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/060,410
; FILING DATE: 14-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 860098.421
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 1:
```

Alignment Scores: 0.0451 Length: 2187  
 Pred. No.: 145.00 Matches: 112  
 Score: 36.04% Conservative: 88  
 Percent Similarity: 36.04% Mismatches: 227  
 Best Local Similarity: 20.18% Indels: 128  
 Query Match: 4.67% Gaps: 20  
 DB: 24

US-09-890-475-1 (1-609) x ABS51736 (1-2187)

```

Qy 34 LysIleValGluThrGluSerThrSerMetAspIleThrIleGlyGlnSerLysGlnPro 53
Db 634 AAACCTTTTCAATAGAGAGAGAGAGAGATGAT-----GAAATCTGAAACAGAA 684
Qy 54 GlnPheLeuLysSerIleAspGluLeuAlaIaPheSerValAlaValGluThrPheLys 73
Db 685 AAACCTTTGGAATACATCGAAGAAATAGTCTTGCTTCAGATCAAGTGGAAAAATACAAG 744
Qy 74 ArgGlnPheAspAspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSer--- 92
Db 745 CTAGATATTGCCAGTTAGAGAAATTTGAAGAGAGAGATGATGAATTTTAAAGCCTT 804
Qy 93 -----LysLeuGluSerAsnGlyValValLeuAla----- 102
Db 805 AAGCAGTCTCTTGAGGAGAAATATTGTATATTCTAAACAAAGTAGAAGATCTAAATGTG 864
Qy 103 -----AlaArgAsnAsnAsnPheHisGln 110
Db 865 AATGTCAGTGTGTAAGAAAGAAAGAACACCATGTCACAGGAATAGAGAACACAC 924
Qy 111 ProMetLeuSerProProArgAsnAsnValSerValGluThrThrValThrValSerGln 130
Db 925 GAAATCTTAAATCAGAGATGCAAACTTA----- 954
Qy 131 ProSerGlnGluLeuValProGluThrSerAsnLysProGluGlyGlyArgMetCysGlu 150
Db 955 -----AAACAGAGAGTTATTCTTGACCAACACAGGAACTG---GAAAGCTTCAACAAAAGAA 1008
Qy 151 LeuMetCysSerLysGlyLeuArgLysIleTyAlaAsnIleSerAspGlnAlaLys 170
Db 1009 TTACAAATGATTCACTCTTCGCAACAGAGAGAAAGATATTCTTCGAGTCTTCATCAGAAG 1068
Qy 171 Leu-----MetGluGluIleProSerAlaLeuLysLeuAlaLysGluProAlaLys 187
Db 1069 CTCTGTCTTTTCAAGAGGAATGGTTAAAGAGAGAAATCTGTTTGAGGAGAAATTAAG 1128
Qy 188 PheValLeuAspCysIleGlyLysPheTyLeu-----GlnGlyArgAlaPhe 204
Db 1129 CAACACTGGATGAGCTTGATTAATACAGCAAAAGGAGGAAACAGCTGAAAGGCTGGTC 1188
Qy 205 ThrLysGluSerProMetSerAlaArgGlnValSerLeuLeuLeuGluSerPhe 224
Db 1189 AAGCAATTGGAAGAGGAAGCAAAATCTAGAGCTGAAGATTTAAACTCTCAGAGAAAG 1248
Qy 225 LeuLeuMetProAspArgGlyLysGlyLysValLysIleGluSerTrpIleLysAspGlu 244
Db 1249 CTG-----AAAGGAAG-----GAGGCTGAA 1269
Qy 245 AlaGluThrAlaAlaValAlaIlePheArgLysArgLeuMetThrGluGlyGlyLeuAla 264
Db 1270 CTGGAGAAAGTAGTGCTGCT-----CATACCAGCCACCCTGCTGCTTTG 1314
Qy 265 AlaGluLysMetAspAla-----ArgGly 272
Db 1315 CAGGAAAGATATGATGATGTCGCAAGCCTTGAGATGTTACTGCTCAATTTGAGGC 1374
Qy 273 LeuLeuLeuValAlaCysPheGlyValProSerAsnPheArgSerThrAspLeuLeu 292
Db 1375 TATAAAGCGTTTAAACAGCCAGTGAGATAGAGATCTTAAAGCTGGAGAACTCATCATACAG 1434
Qy 293 AspLeuLeuArgMetSerGlySerAsn-----GluIleAlaGlyAla 306

```

```

Db 1435 GAAAGCGCGCAAGGCTGGAAAAATGCAGAGATGTTTCAGCATCAGATTTTGGCAACT 1494
Qy 307 LeuLysArgSerGlnPheLeuValProMetValSerGlyIleValGluSerSerIleLys 326
Db 1495 GAGAGCTCAAAATCAAGATATGTAAAGATGCTTTAGATCTGCAGACCAAGTCAGACTA 1554
Qy 327 ArgGlyMetHisIleGluAlaLeuGluMetValTyThrPheGlyMetGluAspLysPhe 346
Db 1555 AAGGAAACAGAAATTAAGAAATCACAGTTCTTTCTTCAAAAAATAACTGAT----- 1608
Qy 347 SerAlaAlaLeuValLeuThrSerPheLeuLysMetSerLysGluSerPheGluArgAla 366
Db 1609 -----TTGCAGACCAACTCAAGCAACAGAGGAGAGACTTTAGAAAAACAG 1653
Qy 367 LysArgLysAlaGlnSerProLeuAlaPheLysGluAlaIaThrLysGlnLeuAlaVal 386
Db 1654 CTGGAAGATAGAGAGAGAGAAAGCTGAAAGAAATACACAGCAGATTA----- 1707
Qy 387 LeuSerSerValMetGlnCysMetGluThrHisLysLeuAspProAlaLysGluLeuPro 406
Db 1708 -----ACTGAAGAAATTAAC 1722
Qy 407 GlyTyrGln-----IleLysGluGlnIleValSerLeuGluLysAsp---ThrLeuGlnLeu 424
Db 1723 AAGTGGCGTCTCTCTATGAAGAACTATATATAAACAACAACCTTTTCAGCTACACTA 1782
Qy 425 AspLys---GluMetGluGluLysAlaArgSerLeuSerLeuMetGluGluAlaAlaLeu 443
Db 1783 GATGCTTTTGAAGTAGAAAAACAGGCA-----TTGTTGAATGAACATGGTGCA 1830
Qy 444 AlaLysArgMetTyrAsn-----GlnGln 451
Db 1831 GCTCAGGAACAGCTAAATAAATAAGAGATTCATATGCTAAATTTATGGGTCAATCAAT 1890
Qy 452 IleLysArgProArgLeuSerProMetGluMetProProValThrSerSerSerTyrSer 471
Db 1891 TTCAACAAATAAATCAAGCATGTTCTGGAAGTTGAAGATGAATAAGTAGCAACTCAATCG 1950
Qy 472 ProIleTyrArgAspArgSerPheProSerGlnArgAspAspGlnAspGluIleSer 491
Db 1951 GAAGTATCAAAACTCCGCTGTCAGCTTCTTAAAAAACAACAAAGTGAGACAAAACCTCAA 2010
Qy 492 AlaLeuValSerSerTyrLeuGly-----ProSerThrSerPheProHis 506
Db 2011 GAGGAATTGAATAAAGTTCTAGGTATCAACACATTGATCTTCAAGGCTTTTTCATCAT 2070
Qy 507 ArgSerArgArgSerProGluTyrMetValProLeuProHisGly 521
Db 2071 GAAAGTAAAGAAAAATTTTGGCCCTGAAGACCCCATTAAGAAAGGC 2115

```

Search completed: August 25, 2003, 10:39:46  
 Job time : 547 secs

Db 1459 CAGAGGGCTGCCATCTG-----CAGACTGAGGTGGATGCT 1494  
QY 116 ProArgAsnValSerValGluThrThrValThrValSerGlnProSerGlnGluIle 135  
Db 1495 CTCGGATTGCGTTTGAAGAGAGAAACCATG--TTGAATAAAAGACAAACCAATT 1551  
QY 136 ValProGluThrSerAsnLysProGluGlyArgMetCysGluLeuMetCysSerLys 155  
Db 1552 CAGGATATGGCTGAAGAGAG-----GGGACACAGCTGGAGAGATACATGACCTCAAG 1605  
QY 156 GlyLeuArgLysTyrIleTyrAlaAlaSerAspGlnAlaLysLeuMetGluGluIle 175  
Db 1606 GACATGTTGGATGTCAGAGAGCGGAAGTTAATGTTCTTCAGAGAAGATTGAATATCTT 1665  
QY 176 ProSerAlaLeuLysLeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLys 195  
Db 1666 CAAGAGCAGCTTAGA----- 1680  
QY 196 PheTyrLeuGlnGlyArgAlaPheThrLysGluSerProMetSerSerAlaArgGln 215  
Db 1681 -----CACAGGAAGAGCAGATGACGACTTGAAGAA 1713  
QY 216 -----ValSerLeuLeuLeuLeuGlu 222  
Db 1714 CGGGTCAATCTCCAGCGCTGACACCACCACTGACACTGCGCTTGACAACTTTGGAG 1773  
QY 223 SerPheLeuLeuMetProAspArgGlyLysGlyLysValLysIleGluSerTyrPheLys 242  
Db 1774 GAGGCCCTTGCAGAAAGAGCGGCAATTGAACGCTTAAG---GAGCAGAGGGACAGA 1830  
QY 243 AspGluAlaGluThrAlaAlaValAlaTyrArgLysArgLeuMetThrGluGlyLeu 262  
Db 1831 GATGAGCGAGAG----- 1842  
QY 263 AlaAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuValAlaCysPheGlyVal 282  
Db 1843 ----AGCAAGAGGAATTCAT----- 1860  
QY 283 ProSerAsnPheArgSerThrAspLeuLeuAsp-----LeuIleArg 296  
Db 1861 -----AACTACAAAAA---GATCTTAAGACTTGAAGAAAAAGTCAGCGTGTTCGAA 1911  
QY 297 MetSerGlySerAsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMet 316  
Db 1912 GCGACCTTTCAGAAAGAGAGCTTCATCTTTGGAT-----CTGAAGAGCAT 1959  
QY 317 ValSerGlyIleValGluSerSerIleLysArgGlyMetHisIleGluAlaLeuGluMet 336  
Db 1960 GCTTCTCTCGCATCCTCAGAGCTGAAAAGGACTCAGCGCTTAAGACACTAGAGATT 2019  
QY 337 ValTyrThrPheGlyMetGluAspLysPheSerAlaAlaLeuValLeuThrSerPheLeu 356  
Db 2020 -----GCTTTGGAGCAGAAAGAGGAGGAGTGTCTGAAAATGGAATCACAATTG 2067  
QY 357 Lys-----MetSerLysGluSerPheGluArgAla 366  
Db 2068 AAAAAGGCACATGAGCGCAGCATTGAAGCCAGCCAGCTCAGAGATGATGACCGAATA 2127  
QY 367 LysArgLysAlaGlnSerProLeuAlaPheLysGluAlaAlaThrLysGlnLeuAlaVal 386  
Db 2128 CAGCAGCTTGGAGAGAGAGATCACCAGGTACAAAGATGAATCTAGCAGGCCAGCGCAGAA 2187  
QY 387 LeuSerSerValMetGluCysMet-----GluThrHisLysLeuAspProAlaLys 403  
Db 2188 GTTATCGACTTTAGAAATCTTTGAGAGAGGTGGAATGATGAAGAATGAAGAATGAATG 2247  
QY 404 Glu-----LeuProGlyTyrGlnLysGluGln-----Ile 414  
Db 2248 AAGATAGCTGAGTTGGAAGTCTCAGCTCAAGCAAGTGAAGAACAGCAAGTAAGAAGTA 2307  
QY 415 ValSerLeuGluLysAspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSer 434  
Db 2308 GCAAAATCTGAAG-----CACAGGAACAGGTGGGAAAAAAGAGAGAT 2349

QY 435 LeuSerLeuMetGluGluAla 441  
Db 2350 GCACAAATGTTAGAGGAGCG 2370  
RESULT 15  
ABS51736  
ID ABS51736 standard; cDNA; 2187 BP.  
XX  
AC ABS51736;  
XX  
DT 21-OCT-2002 (first entry)  
XX  
DE Human novel polynucleotide #9.  
XX  
KW Human; gene; as; trauma, viral infection; parasitic infection; addiction;  
KW bacterial infection; Alzheimer's disease; Huntington's disease; anxiety;  
KW Parkinson's disease; behavioural disorder; pain; hair growth disease;  
KW alopecia; pigmentation disorder; inflammatory disorder; arthritis; AIDS;  
KW inflammatory bowel disease; Crohn's disease; cancer; adenocarcinoma;  
KW acquired immunodeficiency syndrome; colon; asthma; hypertension; obesity;  
KW autoimmune disease; diabetes; graft versus host disease; ulcer; bulimia;  
KW anorexia; dementia; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN WO200257452-A2.  
XX  
PD 25-JUL-2002.  
XX  
PF 17-DEC-2001; 2001WO-US49122.  
XX  
PR 15-DEC-2000; 2000US-256025P.  
PR 30-JAN-2001; 2001US-265163P.  
PR 02-MAR-2001; 2001US-272929P.  
PR 09-MAR-2001; 2001US-274864P.  
PR 16-MAR-2001; 2001US-276888P.  
PR 22-MAR-2001; 2001US-277880P.  
PR 25-APR-2001; 2001US-286409P.  
PR 31-JUL-2001; 2001US-309246P.  
PR 29-AUG-2001; 2001US-315600P.  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
PI Shinkets RA, Colman SD, Spytek KA, Ballinger RA, Guo X;  
PI Tchernev VT, Shenoy SG, Li L, Ellerman KE, Zerhusen BD;  
PI Patturajan M, Casman SJ, Boldog F, Gusev VV, Burgess CE;  
PI Edinger S, Gangolli EA, Malyankar UM, Gunther E, Smithson G;  
PI Millet I, Gerlach VL;  
XX  
DR WPI; 2002-590743/63.  
XX  
P-PSDB; ABG70293.  
XX  
PT Novel polypeptide, designated NOVX for treating or preventing disorders  
PT or symptoms e.g. trauma, Alzheimer's disease, cancers, acquired  
PT immunodeficiency syndrome, asthma and rheumatoid arthritis -  
XX  
PS Claim 9; Page 47; 252pp; English.  
XX  
CC The invention relates to human novel polynucleotides and polypeptides.  
CC The sequences are useful for the treatment, prevention and diagnosis of  
CC disorders such as trauma, viral/parasitic/bacterial infections,  
CC Alzheimer's disease, Huntington's disease, Parkinson's disease,  
CC behavioural disorders, anxiety, addiction, pain, hair growth diseases,  
CC alopecia, pigmentation disorder, inflammatory disorders such as osteo-  
CC and rheumatoid arthritis, inflammatory bowel disease, Crohn's disease,  
CC acquired immunodeficiency syndrome (AIDS), cancers such as colon cancer  
CC and adenocarcinoma, asthma, hypertension, autoimmune disease, diabetes,  
CC obesity, graft versus host disease, ulcer, bulimia, anorexia and  
CC dementia. Sequences ABS51728-ABS51748 represent human novel  
CC polynucleotides of the invention.  
XX  
SQ Sequence 2187 BP; 872 A; 354 C; 461 G; 500 T; 0 other;

Db 2513 GGAGAACAGACCCGGAATACCTATCTTGGCTGAGCAGTATGATCAGACATTATGA 2572  
 Qy 392 nCysMetGluThrHis-----LysLeuAspProAlaLysGluLeuProGlyTyrGlnIle 410  
 Db 2573 AATGCTCTCCACACAAAGCCCTGCTTGGTTGGATGAAGCAGACAGAGAGTCCAGGTTT 2632  
 Qy 410 eLysGluGlnIle-----ValSerLe 417  
 Db 2633 GAAGATGCAGCTCAGCAGGAACTGGAGCTGTTGAATGCGTATCAGAGCAAAATCAAGAT 2692  
 Qy 417 uGluLysAspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSerLeuSerLe 437  
 Db 2693 GCAAGCTGAGGCA---CAACATGATCGAGCTTCGCGAGCTTGAACAGAGGGTCTCCCT 2749  
 Qy 437 u-----MetGluGluAlaLeuAlaLysArgMetTyrAs 449  
 Db 2750 CCGAGGGCACTTTAGAACAAAGATTGAAGAGATGTTGGT-----TTCAGAA 2803  
 Qy 449 nGlnGlnIleLysArgProArgLeuSerProMetGluMetProProValThrSerSerSe 469  
 Db 2804 TGAGCGCAGACGAATACGA----- 2825  
 Qy 469 rTyrSerProIleTyrArgAspArgSerPheProSerGlnArgAspAspGlnAspG 489  
 Db 2826 -----AGCTGTGGAACTCAAGCCAGCAGGA 2851  
 Qy 489 uIleSerAlaLeuValSerSerTyrLeuGlyProSerThrSerPheProHisArgSerAr 509  
 Db 2852 GAUTGAAGCTTTTGACTCTGAAGACATGAGACTAGGTTTATGATATGCTCTCTTCAA 2911  
 Qy 509 gArgSerProGluTyrMetValProLeuProHisGlyGlyLeuGlyArgSerValTyrAl 529  
 Db 2912 TCTCTCCCTCGAG----- 2926  
 Qy 529 aTyrGluHisLeuAlaProAsn-----SerTyrSerProGlyHisGly----- 543  
 Db 2927 ATTCAAGCCAGCAGCTACCGGGAGCTTCTGTTGTCACACAACTTACTGGGGGTCAGG 2986  
 Qy 544 ----HisArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGlnArgHisProLe 562  
 Db 2987 ACCTCACTGGGGTCATCCCATGGTGGCCACACACAGCTTGGGGC-----CAPCCAA 3040  
 Qy 562 uGlnTyrSerPro-----ProIleHisGlyGlnGlnGlnLeuProTyrG 577  
 Db 3041 GCAAGTGGACCCAGCCATCGGCTGACCTTCAGGGCCCAATGCAAGGGGTACTTCGAGG 3100  
 Qy 577 yIleGlnArgValTyrArgHisSerProSerGluGluArgTyrLeu-----GlyLe 594  
 Db 3101 TAGCAGTATGGAGTCCGCAATAGCCCGGACAGCTCTGAGGCGGACAGCTTCTGGGGGAGC 3160  
 Qy 594 uSerAsnGlnArgSerProArgSerAsnSer 604  
 Db 3161 GACGAGCAGGCGATGAGCAGACGAGT 3191  
 RESULT 14  
 ABK32901  
 ID ABK32901 standard; DNA; 2760 BP.  
 AC ABK32901;  
 DT 23-APR-2002 (first entry)  
 DE DNA encoding human homologue of SRB4 antifungal target.  
 KW antifungal; fungal gene transcription; RPC34; POP3; TPA2; NAB2;  
 KW MPT1; MTR2; BOS1; POL30; RSA2; SOT1; MTW1; TPB1; SPC98; BFR2; RNA1;  
 KW GCD7; SK16; NIP1; LCP5; NCE103; ECO1; ORC2; CNS1; YPD1; TIM10; SRB4;  
 KW yeast; fungus; ds; gene.  
 OS Homo sapiens.  
 XX W0200202055-A2.  
 XX

PD 10-JAN-2002.  
 XX 28-JUN-2001; 2001WO-US20592.  
 DF 29-JUN-2000; 2000US-215164P.  
 PR 10-AUG-2000; 2000US-224457P.  
 XX (ANAD-) ANADYS PHARM INC.  
 PA Moore J, Buurman ET, Desilva T, Harris S, Komarnitsky S;  
 PI Mendillo M, Moore D, McCoy M, Sanderson K, Haq T, Zhu S, Long F;  
 PI Davidov E, Thompson CM;  
 XX WPI; 2002-147962/19.  
 DR P-PSDB; AAU83013.  
 XX Screening candidate antifungal compound for interaction with essential  
 PT protein, modulation of essential protein activity, binding to essential  
 PT protein, by contacting protein with test compound and determining  
 PT effects  
 XX Disclosure; Figure 80; 522pp; English.  
 PS The invention describes a method of screening a candidate antifungal  
 CC compound for interaction with essential proteins (EP) or for modulation  
 CC of EP activity e.g. fungal gene transcription. The proteins tested in the  
 CC invention include RPC34, POP3, TPA2, NAB2, MPT1, MIR2, BOS1, POL30, RSA2,  
 CC SOT1, MTW1, TPB1, SPC98, BFR2, RNA1, GCD7, SK16, NIP1, LCP5, NCE103,  
 CC ECO1, ORC2, CNS1, YPD1, TIM10 and SRB4 from *S. cerevisiae*, *C. albicans*  
 CC and human homologues. The method involves contacting a culture with one  
 CC or more test compounds and determining the effects on the growth or  
 CC viability of the culture of cells which preferably comprises fungal cells  
 CC or yeast cells. Preferably the identified compounds interact with or  
 CC modulate (preferably inhibit) activity of *C. albicans* EP. The inhibitor  
 CC compounds identified by the method are useful for preventing or  
 CC inhibiting fungal, particularly *C. albicans* growth in culture or in a  
 CC mammal. The antifungal agents interact with essential fungal elements  
 CC that can be used to treat fungal infection by preventing the growth and  
 CC preferentially killing the fungi, but does not inhibit the biological  
 CC activity of mammalian homologues. This sequence encodes a target protein  
 CC used to test the antifungal compounds, described in the method of the  
 CC invention.  
 XX Sequence 2760 BP; 923 A; 537 C; 765 G; 535 T; 0 other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 0.0565 Length: 2760  
 Score: 145.50 Matches: 95  
 Percent Similarity: 37.69% Conservative: 81  
 Best Local Similarity: 20.34% Mismatches: 164  
 Query Match: 4.69% Indels: 127  
 DB: 24 Gaps: 20  
 US-09-890-475-1 (1-609) x ABK32901 (1-2760)  
 Qy 16 AlaAsnProLeuLeuGlnArgHisGlnSerGluGlnArgArgGluLeuProLysIle 35  
 Db 1228 TCGAATGCTGCTTTCAGTACTGAGCAAGAGGAGAAATGAAAGCAATG---GAAGTG 1284  
 Qy 36 ValGluThrGluSerThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPhe 55  
 Db 1285 TATCGGAGCCATCTTAATTTATGAAATAAGATTGCCAGGTGGAACAG---GAGCTG 1341  
 Qy 56 LeuLysSerIleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGln 75  
 Db 1342 TCCAGAAAGGACACAGAACTACTCCCTCGCAGACAAAAGCTAGAAACACTCACAACACAG 1401  
 Qy 76 PheAspAspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGlu 95  
 Db 1402 TTTCCAGATAGTAAACAGCACATTGAGTGTGAGGAGTCTTGAAGCTCTTGAAGCTCTTGAAG 1458  
 Qy 96 SerAsnGlyValValLeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSerPro 115  
 XX

28-JUN-2002 (first entry)  
 Novel human coding sequence SEQ ID NO: 150.  
 Human; antianemic; vulnerary; antiinflammatory; immunomodulator;  
 antinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;  
 neuroprotective; antiparkinsonian; protein therapy; EST;  
 expressed sequence tag; gene; ss.  
 Homo sapiens.  
 WO200222660-A2.  
 21-MAR-2002.  
 10-SEP-2001; 2001WO-US26015.  
 11-SEP-2000; 2000US-0659671.  
 (HYSE-) HYSEQ INC.  
 Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
 Xue AJ, Yang Y, Wehrman T, Drmanac RT;  
 WI; 2002-292408/33.  
 DR P-PSDB; ABB97326.  
 An isolated polynucleotide for treating diseases associated with its  
 encoded polypeptide such as cancer and multiple sclerosis -  
 Claim 1; SEQ ID NO 150; 509pp; English.  
 The present invention provides the protein and coding sequences of 444  
 novel human proteins. These were isolated from expressed sequences tags  
 (ESTs). They can be used to stimulate cell growth, to regulate  
 haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth  
 e.g. in burn treatment, to regulate the immune system e.g. to treat  
 multiple sclerosis, to regulate activin or inhibin e.g. to treat  
 infertility, to regulate haemostasis or thrombolysis e.g. to treat  
 stroke and cancer, to screen for drugs, to treat inflammatory conditions  
 e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.  
 Parkinson's disease. The present sequence is a coding sequence of the  
 invention.  
 XX  
 SQ Sequence 4972 BP; 1579 A; 1001 C; 1127 G; 1263 T; 2 other;  
 Alignment Scores:  
 Pred. No.: 0.0719 Length: 4972  
 Score: 148.50 Matches: 150  
 Percent Similarity: 34.88% Conservative: 98  
 Best Local Similarity: 21.10% Mismatches: 227  
 Query Match: 4.78% Indels: 237  
 DB: 24 Gaps: 35  
 JS-09-890-475-1 (1-609) x ABB59739 (1-4972)  
 2Y 3 AsnThr-----ProProThrValAlaAlaGlnProThrThrAlaAsn 17  
 DE 1443 RAATACAGAGAGAGGAGATCCTAGACACAGAGCATCAGATCCCAATCTCCA----- 1496  
 2Y 18 ProLeuLeuGlnArgHisGlnSerGluGlnArgArgGluLeuProLysIleValGlu 37  
 DB 1497 CCCCAGATATCTCGTCAAAATCAGATATCGTAAATCGAGACACTTTGTACTATACGG 1556  
 2Y 38 ThrGluSerThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLys 57  
 DB 1557 ACAGATCATCTG-----GTTACGAGGCAAAATGCAAGAACATGAG 1595  
 2Y 58 SerIleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAsp 77  
 DB 1596 CAGGACTCTGAGCTTAGAGACAAATGCTCGCTATAGCGAATGAGGCGCAACAT--- 1652  
 2Y 78 AspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsn 97

Qy 18 ProLeuLeuGlnArgHisGlnSerGluGlnArgArgGluLeuProLysIleValGlu 37  
Db 1419 CCCAAGTATCTCGTACAAAATCACACTATCGTAATCGAGAACACTTTGCTACTATACGG 1478  
Qy 38 ThrGluSerThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLys 57  
Db 1479 ACAGCATCTG-----GTTACGAGCGCAAAATCAAGAACATGAG 1517  
Qy 58 SerIleAspGluLeuAlaPheSerValAlaValGluThrPheLysArgGlnPheAsp 77  
Db 1518 CAGGACTCTGAGCTTAGAGAACAAATGCTGCTGATATAAGCGAATGAGCGCAACAT--- 1574  
Qy 78 AspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGlnSerAsn 97  
Db 1575 -----CAAAAGCACTGATGCTCTGGAACAAAGCTAAAGCTGAGATGATGAACAT 1628  
Qy 98 GlyValValLeuAlaAlaArgAsnAsnPheHisGlnProMetLeuSerProArg 117  
Db 1629 CGCTCAGATTA-----GACAAAGATCTTGAAACTCAGCGT 1664  
Qy 118 AsnAsnValSerValGluThrThrValThrValSerGlnProSerGlnGluIleValPro 137  
Db 1665 AACAAATTTCTGTCAGAAATGAGAAACTTATCAAGAAACACAGGCTGCTATGAGAAA 1724  
Qy 138 GluThr-----SerAsnLysProGluGlyGlyArgMetCysGluLeuMetCysSer 154  
Db 1725 GAGGCTAAAGTGATGTCCATGAA-----GAG 1751  
Qy 155 LysGlyLeuArgLysIleValAlaAsnIleSerAspGlnAlaLysLeuMetGluGlu 174  
Db 1752 AAAAAATTCAGAACATATT-----CAGCCCAACAGAAAGAA 1793  
Qy 175 IleProSerAlaLeu-----LysLeuAlaLysGluProAlaLys 187  
Db 1794 CTGAATAGTTTCTCGAGTCCCAAGAAAGAGAGATATAAACTTCGAAAGAGCAGCTTAA 1853  
Qy 188 PheValLeuAspCysIleGlyLysPheThrLeuGlnGlyArg-----AlaPhe 204  
Db 1854 GAGGAGCTAAAT---GAAACCCAGAG-TACCCCCAAAGAAACACAGAGTGCTTTC 1909  
Qy 205 ThrLysGluSer-----ProMetSerSer 212  
Db 1910 AAGCAGAGGAGGAATATACAGCATTTCCAAAGCAGAAAGAGAGTAACTCTCTTCGAGC 1969  
Qy 213 AlaArgGlnValSerLeuLeuIleLeuGluSerPheLeuLeuMetProAspArgGlyLys 232  
Db 1970 TCAAGACAATA-----CCTAGAGCTGGAATG 1996  
Qy 233 GlyLysValLysIleGluSerTrpIleLysAspGluAlaGluThrAlaAlaValAlaTrp 252  
Db 1997 CCCTCGCTTCAAGAGAGAAAT-----GTTACTTGG 2026  
Qy 253 ArgLysArgLeuMetThr-----GluGlyGlyLeuAlaAlaAlaGluLysMetAsp 269  
Db 2027 CGCTCATAACTTAGACGAGGACCTTGTGAGGAGAGATT-----AAACAAAGACAGAC 2080  
Qy 270 AlaArgGlyLeuLeuLeuValAlaCysPheGlyValProSer----- 284  
Db 2081 TCAGAGAGACTAGA-----GCATGCCATGCTACTCCGACAGCATGATCTATGCA 2131  
Qy 285 -----AsnPheArgSerThrAspLeuLeu-AspIleuIleArgMe 297  
Db 2132 AGAACTGGAGTTCCGCCACCTCAACACAATTCAGAGATGCGTGTGAGTGTATCAGATT 2191  
Qy 297 tSerGlySerAsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetVa 317  
Db 2192 ACAGATCAAACTGAGCTACTACACAGCTGGAATATAATAAGCGAAGA----- 2240  
Qy 317 lserGlyIleValGluSerSerIleLysArgGlyMetHisIleGluAlaLeuMetVa 337  
Db 2241 -----GAACGAGAACTAAGACGA---AAGCATGTC----- 2267  
Qy 337 lTyrThrPheGlyMetGluAspLysPheSerAlaAlaLeuValLeuThrSerPheLeuLys 357

Db 2268 -----ATGGAAGTTTCGACCAACAGCCTTAGAGAGTTTGAAGTCTTAAGAACTCCA 2314  
Qy 357 sMetSerLysGluSerPheGluArgAlaLysArgLysAlaGlnSerProLeuAla----- 375  
Db 2315 AATAAAAAGCAGTTTTCAGGATACCTCGAAATAATCCAAACACAGACAGTACAAAGCATTAAG 2374  
Qy 376 -----PheLys 377  
Db 2375 AAATCACCTGCTGGAGACTACACCAAAAGAGTGACACAAAGCTGTTCTCGAAACGGCTCAA 2434  
Qy 377 sGluAlaAlaThrLysGlnLeuAlaValLeuSer-----SerValMetG1 392  
Db 2435 CGAGAAACAGACCCGGAATTAGCTATCTTGGCTGAGCAGTATGATCAGACCATTAATGA 2494  
Qy 392 nCysMetGluThrHis-----LysLeuAspProAlaLysGluLeuProGlyTrpGlnI1 410  
Db 2495 AATGCTCTCCACACAAAGCCCTGCGTTTGGATGAAGCACAGGAAGCAGAGTCCAGGTTTT 2554  
Qy 410 eLysGluGlnIle-----ValSerLe 417  
Db 2555 GAAGATGCACTGTCAGCAGGAACCTGAGCTGTTGAATCGGTATCAGAGCAAAATCAAGAT 2614  
Qy 417 uGluLysAspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSerLeuSerLe 437  
Db 2615 GCAAGCTGAGGCA---CAACATGATCGAGAGCTTCGCGAGCTTCGACAGAGGGTCTCCCT 2671  
Qy 437 u-----MetGluGluAlaAlaLeuAlaLysArgMetTyrAs 449  
Db 2672 CCGAGGGCACTCTTAGAACAAAGATTGAAGAGAGATGTTGGCT-----TTGCAGAA 2725  
Qy 449 nGlnGlnIleLysArgProArgLeuSerProMetGluMetProValThrSerSerSe 469  
Db 2726 TGAGCCACAGAACGAAATACGA----- 2747  
Qy 469 rTyrSerProIleTyrArgAspArgSerPheProSerGlnArgAspAspGlnAspG1 489  
Db 2748 -----AGCCTGTTGGAACGTCACAGCCACAGA 2773  
Qy 489 uIleSerAlaLeuValSerSerTyrLeuGlyProSerThrSerPheProHisArgSerAr 509  
Db 2774 GATTGAAGTTTTTGACTCTGAAGCATGAGACTAGGTTTTAGTAATATGTCCTTTCTAA 2833  
Qy 509 gArgSerProGluTyrMetValProLeuProHisGlyLeuGlyArgSerValTyrAl 529  
Db 2834 TCTCTCCCTGAG-----GC 2848  
Qy 529 aTyrGluHisLeuAlaProAsn-----SerTyrSerProGlyHisGly----- 543  
Db 2849 ATTCAGCCACAGCTACCCGGAGCTTCGGTGTGTCACACACACCTACTGGGGTCCAGG 2908  
Qy 544 ----HisArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGlnArgHisProLe 562  
Db 2909 ACCTCACTGGGTTCATCCCATGGGTGGCCACCACCAAGCTTGGGC-----CATCCAAT 2962  
Qy 562 uGlnTyrSerPro-----ProIleHisGlyGlnGlnLeuProTyrG1 577  
Db 2963 GCAAGTGGACCCCAACCCATCGGCTCACCTTCAGGGCCCAATGCAAGGGTACCTCCGAGG 3022  
Qy 577 yIleGlnArgValTyrArgHisSerProSerGluGluArgTyrLeu-----GlyLe 594  
Db 3023 TAGCAGTATGGGAGTCCGAATAGCCCCCAGGCTCTGAGCGGACAGCTTCTGGGGGACG 3082  
Qy 594 uSerAsnGlnArgSerProArgSerAsnSer 604  
Db 3083 GACAGACGAGGATGAGCAGAGCAGGAT 3113  
RESULT 13  
AENS9739 standard; cDNA; 4972 BP.  
ID ABNS9739  
XX  
AC ABNS9739;  
XX

QY 317 lserGlylleValGluSerSerIleLysArgGlyMetHisIleClnAlaLeuGluMetVa 337  
Db 1445 -----GACGAGAACTAAGACGA---AAGCATGTC----- 1471  
QY 337 lTyrThrPheGlyMetGluAplysPheSerAlaAlaLeuValLeuThrSerPheLeuLy 357  
Db 1472 -----ATGGAAGTTCGACCAACAGCCCTAAGAGTTTGAAGCTCAAAAGAACTCCA 1518  
QY 357 smetSerLysGluSerPheGluArgAlaLysArgLysAlaGlnSerProLeuAla----- 375  
Db 1519 AATAAAGAGAGTTTCAGGATACCTCGCAAAATCCAAACAGCAGATACAAAGCATTAAG 1578  
QY 376 -----PheLy 377  
Db 1579 AAATCACCTGTGGAGACTACACAAAGAGTGAGCACAAAGCTTCTGAAACGGCTCAA 1638  
QY 377 sClnAlaAlaThrLysGlnLeuAlaValLeuSer-----SerValMetGln 392  
Db 1639 GGAGGAACAGACCCCGAAATTAGCTATCTTGCTGAGCAGTATGATCACAGCATTAATGA 1698  
QY 392 nCysMetGluThrHis-----LysLeuAspProAlaLysGluLeuProGlyTrpGlnIl 410  
Db 1699 ATGTCTCTCCACACAGCCCTGCGTTTGGATGAGCACAGGACGAGTGCAGGTTT 1758  
QY 410 elysGluGlnIle-----ValSerLe 417  
Db 1759 GAAGATGCAGCTGCAGCAGGAAGTGGAGCTGTGAATCGCTATCAGAGCAAAATCAAGAT 1818  
QY 417 uClnLysAspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSerLeuSerLe 437  
Db 1819 GCAAGCTGAGGCA---CAACATGATCGAGAGCTTCGGAGCTTGAACAGAGGTCTCCCT 1875  
QY 437 u-----MetGluGluAlaAlaLeuAlaLysArgMetTyrAs 449  
Db 1876 CCGGAGGGCAGCTCTTAGAACAAAAGATTGAAGAAGAGATGTGGCT-----TTGCAGA 1929  
QY 449 nClnGlnLysArgProArgLeuSerProMetGluMetProValThrSerSerSe 469  
Db 1930 TGAGGCGCAGAAACAAATACGA----- 1951  
QY 469 rTyrSerProIleTyrArgAspArgSerPheProSerGlnArgAspAspGlnAspGl 489  
Db 1952 -----AGCTGTGGAAAGTCAAGCCAGAGA 1977  
QY 489 ulleSerAlaLeuValSerSerTyrLeuGlyProSerThrSerPheProHisArgSerAr 509  
Db 1978 GATTGAAGCTTTTGACTCTGAAGCATGAGACTAGGTTTAGTAAATATGTCCTTTCTAA 2037  
QY 509 gArgSerProGluTyrMetValProLeuProHisGlyGlyLeuGlyArgSerValTyrAl 529  
Db 2038 TCTCTCCCTGAG-----GC 2052  
QY 529 aTyrGluHisLeuAlaProAsn-----SerTyrSerProGlyHisGly----- 543  
Db 2053 ATTACGCCACAGCTACCCGGGAGCTCTGGTTGGTCAACACACCTACTCGGGGTCCAGG 2112  
QY 544 ----HisArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGlnArgHisProLe 562  
Db 2113 ACTCACTGGGTATCCATGGTGGGCCACCAAGCTTGGGC-----CATCCAT 2166  
QY 562 uGlnTyrSerPro-----ProIleHisGlyGlnGlnLeuProTyrGln 577  
Db 2167 GCAAGGTGGACCCAGCCATGGGGTCAACCTTCAGGGCCAAATGCAAGGGGTACTCGAGG 2226  
QY 577 yIleGlnArgValTyrArgHisSerProSerGluGluArgTyrLeu-----GlyLe 594  
Db 2227 TAGCAGTATGGAGTCCGCAATAGCCCCCAGGCTCTGAGGCGGACAGCTTCTGGGGGAGC 2286  
QY 594 uSerAsnGlnArgSerProArgSerAsnSer 604  
Db 2287 GACGGAGCAGGGCATGAGCAGAGAACCAAGT 2317

RESULT 12  
ACC50205  
ID ACC50205 standard; cDNA; 4620 BP.  
XX  
AC ACC50205;  
XX  
DT 12-JUN-2003 (first entry)  
XX  
DE Breast cancer associated cDNA sequence SEQ ID NO:253.  
XX  
KW Human; breast cancer; cytostatic; gene therapy; gene; ss.  
XX  
OS Homo sapiens.  
XX  
FN WO2003004989-A2.  
XX  
PD 16-JAN-2003.  
XX  
PF 21-JUN-2002; 2002WO-US19669.  
XX  
PR 21-JUN-2001; 2001US-299887P.  
XX  
PR 27-JUN-2001; 2001US-301572P.  
XX  
PR 18-JUL-2001; 2001US-306501P.  
XX  
PR 25-SEP-2001; 2001US-325002P.  
XX  
PR 05-MAR-2002; 2002US-362585P.  
XX  
PR 14-MAY-2002; 2002US-380391P.  
XX  
PA (MILL-) MILLENIUM PHARM INC.  
XX  
PI Lillie J, Gannavazapu M, Glatt K, Hoersh S, Kamatkar S, Mertens M;  
PI Monahan JE, Myer V, Wang Y, Xu Y, Zhao X, Meyers RE, Bast RC;  
PI Hortobagyi GN, Pusztai L, Meric F, Sahin A, Mills GB;  
XX  
DR WPI; 2003-210381/20.  
XX  
P-PSDB; ABR47509.  
PT Breast cancer diagnosis or treatment by comparing the level of  
PT expression of a marker in a patient sample with that in the control  
XX non-breast cancer sample  
PS Claim 1; SEQ ID 253; 128pp; English.  
XX  
CC The present invention describes a method for assessing whether a patient  
CC is afflicted with breast cancer. The method comprises comparing the level  
CC of expression of a marker (gene/polypeptide see ACC50076 to ACC50334 and  
CC ABR47386 to ABR47632) in a patient sample and the normal level of  
CC expression of the marker in a control non-breast cancer sample, where a  
CC significant increase in the level of expression of the marker in the  
CC patient sample and the normal level is an indication that the patient is  
CC afflicted with breast cancer. The breast cancer associated sequences  
CC from the present invention have cytostatic activities and can be used in  
CC gene therapy. The method is useful for diagnosing and treating breast  
CC cancer.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 4620 BP; 1490 A; 942 C; 1039 G; 1149 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 0.065 Length: 4620  
Score: 148.50 Matches: 150  
Percent Similarity: 34.88% Conservative: 98  
Best Local Similarity: 21.10% Mismatches: 227  
Query Match: 4.78% Indels: 237  
DB: 25 Gaps: 35  
  
US-09-890-475-1 (1-609) x ACC50205 (1-4620)  
  
QY 3 AsnTyr-----ProProThrValAlaAlaGlnProThrThrAlaAsn 17  
Db 1365 AATTACAGAGAAGAGGGAGATCCTTAGAACAGAGCATCAGATCCCAATCTCCA----- 1418

AC	AAZ40489;
XX	18-FEB-2000 (first entry)
DT	Human SULU3 DNA.
DE	
XX	Antirheumatic; antiarthritic; antinflammatory; antiallergic; osteopathic;
KW	antipsoriatic; antiarteriosclerotic; antiasthmatic; immunosuppressive;
KW	neuroprotective; cardiant; cerebroprotective; cytosstatic; antidiabetic;
KW	vulnery; SRE20; protein kinase; STLK2; STLK3; STLK4; STLK5; STLK6; STLK7;
KW	ZC1; ZC2; ZC3; ZC4; KHS2; SULU1; SULU3; GER2; PAK4; PAK5; antagonist;
KW	antibody; gene therapy; rheumatoid arthritis; artherosclerosis; asthma;
KW	inflammatory bowel disease; Crohn's disease; osteoarthritis; psoriasis;
KW	rhinitis; autoimmunity; organ transplantation; multiple sclerosis;
KW	myocardial infarction; cardiovascular disease; stroke; renal failure;
KW	oxidative stress-related neurodegenerative disorder; Parkinson's disease;
KW	amyotrophic lateral sclerosis; Leigh syndrome; cancer; cardiomyopathy;
KW	ischemic disorder; inflammation; diabetes mellitus; fibrosis; micosis;
KW	mesangial disorder; growth regulation; wound healing; T cell activation;
KW	immunosuppressant; ss.
XX	Homo sapiens.
OS	
XX	WO953036-A2.
PX	21-OCT-1999.
PD	
XX	13-APR-1999; 9SW-US08150.
PF	
PP	
PR	14-APR-1998; 98US-0081784.
PR	(SUCG-) SUGEN INC.
PA	
XX	Plowman G, Martinez R, Whyte D;
PI	WI; 1999-611301/52.
DR	P-PSDB; AAY55937.
DR	
PT	Novel kinase-related polypeptides used for the diagnosis and treatment
PT	of kinase-related diseases and disorders -
PT	P-PSDB; AAY55937.
PS	Disclosure; Page 290-292; 387pp; English.
XX	This sequence represents the coding sequence for a novel STE20-related
CC	protein kinase. The invention relates to nucleic acid molecule encoding
CC	a kinase polypeptide selected from STLK2, STLK3, STLK4, STLK5, STLK6,
CC	STLK7, ZC1, ZC2, ZC3, ZC4, KHS2, SULU1, SULU3, GER2, PAK4 and PAK5. The
CC	proteins are used to identify agonists and antagonists, and to raise
CC	antibodies. The polynucleotides are useful in gene therapy protocols. The
CC	polynucleotides, polypeptides, antibodies, antagonists and agonists may
CC	be used to treat diseases such as immune-related disorders and diseases
CC	(e.g. rheumatoid arthritis, artherosclerosis, asthma, osteoarthritis, psoriasis,
CC	disease (e.g. Crohn's disease), amyotrophy, multiple sclerosis, organ
CC	chronic inflammatory pelvic disease, multiple sclerosis, organ
CC	transplantation, myocardial infarction, cardiovascular disease, stroke,
CC	renal failure, oxidative stress-related neurodegenerative disorders (e.g.
CC	amyotrophic lateral sclerosis, Parkinson's disease and Leigh syndrome),
CC	cancer, cardiomyopathies, ischemic disorders, inflammatory disorders,
CC	diabetes mellitus, fibrotic and mesangial disorders. The proteins may
CC	also be useful for cell growth regulation (e.g. in wound healing), T cell
CC	activation, mitosis control, and as immunosuppressants.
XX	
SQ	Sequence 3824 BP; 1234 A; 779 C; 858 G; 953 T; 0 other;
XX	
Alignment Scores:	
Pred. No.:	0.0502 Length: 3824
Score:	148.50 Matches: 150
Percent Similarity:	34.88% Conservative: 98
Best Local Similarity:	21.10% Mismatches: 227
Query Match:	4.78% Indels: 237
DB:	20 Gaps: 35

Db 1333 AATTACCAAGAGAGAGATCTCTAGAACAGAGCATCAGCTCCACAGTCTCCA----- 1386  
2Y 18 ProLeuLeuGlnArgHisGlnSerGluGlnArgArgGluLeuProLysLeuValGlu 37  
Db 1387 CCTCAAGTGTCTCGTCAAAATCAATTCGTAATAGAGAACACTTTTCAACTATACGA 1446  
2Y 38 ThrGluSerThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLys 57  
Db 1447 ACAGCATCACTG-----GTTACAGACAGATGCAAGAACATGAG 1485  
2Y 58 SerIleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAsp 77  
Db 1486 CAGGACTCTGAACCTAGAGAACAGATCTCGTTATAAGCGGATGAGGCGACAGCAT--- 1542  
2Y 78 AspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsn 97  
Db 1543 -----CAGAAGCAGCTGATGACTCTCTGAAAAATAAACTGAGGCGCAGAAATGGACGACAT 1596  
2Y 98 GlyVal-----ValLeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeu 113  
Db 1597 CGGCTCAGATTAGACAAAGATCTTTGAAACTCAGCGCAACAATTTTCGCTGCAGAAATG--- 1653  
2Y 114 SerProProArgAsnAsnValSerValGluThrThrValThrValSerGlnProSerGln 133  
Db 1654 -----GAGAACTATTAAAGAAACACCAAGCTTCTATG 1686  
2Y 134 GluIleValProGluThrSerAsnLysProGluGlyArgMetCysGluLeuMetCys 153  
Db 1687 GAAAAAGAGGCTAAAGTGTAGGCGCAACAGGAG----- 1719  
2Y 154 SerLysGlyLeuArgLysTyIleTyIleAlaAsnIleSerAspGlnAlaLysLeuMetGlu 173  
Db 1720 ---AAAAAATCCAAACACACTT-----CAGGCTCAACAGAGAAA 1758  
2Y 174 GluIleProSerAlaLeuLys-LeuAlaLysGluProAlaLysPheValLeuAspCys1 193  
Db 1759 GAACCTGAATAGCTTTTGGAGCTCTCAAAAGAGAGATATAAACTTCGAAAGAGCAGCTT 1818  
2Y 193 eGlyLysPheTyIleGlnGlnArgAlaPheThrLysGluSer-----ProMetSe 211  
Db 1819 AAGGAGGAGCTGAATGAACACAGCAGACACCTTAAAGAAAGAGGAGGAGGATGGCTTTCA 1878  
2Y 211 rSerAlaArgGlnValSerLeuLeuIleLeuGluSerPheLeuLeuMetProAspArg1 231  
Db 1879 AAGCAGAGGAGATATTCAACATTTTCAGCAGACAGAGAGCTAATCTTCTTCGACGT 1938  
2Y 231 YLysGlyLysVal-----LysIleGluSerTrpIleLysAspGluAlaGluTh 247  
Db 1939 CAAGGCGAGTATCTAGAGCTAGAATGTCGCTGCTCA-----AAAGAGAAATGTTACTT 1992  
2Y 247 rAlaAlaValAlaIleAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 267  
Db 1993 GGTCGGCATACTTGAACAGAGCCTTG-----TCAGGAGAGGTT-----AAACAAA 2040  
2Y 267 yMetAspAlaArgGlyLeuLeuLeu-----LeuValAlaC 279  
Db 2041 AGGCAGACTCAGAGAGCTATGAGACATGCAATGTTACTGCGCAGACAGATGAATCCATGCA 2100  
2Y 279 yPheGlyVal---ProSerAsnPheArgSerThrAspLeuLeu-AspLeuIleArgMet 297  
Db 2101 GAACCTGGAGTTTCGCCACTCAACACTATTGAGAGATGGCTGTGAGTTGATCAGACTG 2160  
2Y 298 SerGlySerAsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetVal 317  
Db 2161 CAACATCAAACTGAGCTTACTTAACCCAGCTGGAATACAATAAGAGAGG----- 2208  
2Y 318 SerGlyIleValGluSerSerIleLysArgGlyMetHisIleGluAlaLeuGluMetVal 337  
Db 2209 -----GAACGGGAAGCTAAGACG---AAACATGTC----- 2235  
2Y 338 TyrThrPheGlyMetGluAspLysPheSerAlaAlaLeuValLeuThrSerPheLeuLys 357  
Db 2236 -----ATGGAAGTTCGACAGCAGCCTTAAGAGTTTGAAGTCTTAAGAGACTCCA 2283

QY 358 MetSerLysGluSerPheGluArgAlaLysArgLysAlaGlnSerProLeuAla----- 375  
Db 2284 ATAAAAAGCAGTTTCAGGATACCTGCAAAATTCAAACCCAGACAGTACAAAGCATTAAGG 2343  
QY 376 -----PheLys 377  
Db 2344 AATCACCCTACTGGAGACTACACAAAGAGTGGACACAAAGCTGTTCTGAAAAGACTCAAG 2403  
QY 378 GluAlaAlaThrLysGlnLeuAlaValLeuSer-----SerValMetGln 392  
Db 2404 GAGAACAGACTCGGAAGTTAGCCATCTTTGGCTGAGCAGTATGATCATGATTAATGAA 2463  
QY 393 CysMetGluThrHis-----LysLeuAspProAlaLysGlu----- 404  
Db 2464 ATGCTCTCCACACAAAGCTCGCTGTTGGATGAAGCAGACAGGAAGCAGAAATGCCAGGTTTG 2523  
QY 405 -----LeuProGlyTrpGlnIleLys----- 411  
Db 2524 AAGATGCAAGTACAGCAGGAACTGGAGCTGTTGAATGCATATCAGAGCAAAATCAAGATG 2583  
QY 412 -----GluGlnIleValSerLeu----- 417  
Db 2584 CAGGCTGAGGCCCAACATGATCGAGAGCTTCGAGAGCTGGAACAAAGGGTCTCCCTTCGG 2643  
QY 418 -----GluLysAspThrLeuGlnLeuAspLysGluMetGlu 429  
Db 2644 AGACACTCTTGAACAGAGAGATTGAAGAGAGATGTTGGCTTTGCAGATGAACGCACA 2703  
QY 430 GluLysAlaArgSerLeuSerLeuMetGluAlaAlaLeuAlaLysArgMetTyrAsn 449  
Db 2704 GAACGAATACGT-----AGCCTGCTCGAGCGCCAGAGCCAGAGAAATTAAGCTTTTGAC 2757  
QY 450 GlnGlnIleLysArgProArgLeuSerProMetGluMetProProValThrSerSerSer 469  
Db 2758 TCTGAAAGCATGAGATAGGTTTATGTAACATGCTCTT-----TCTAAT 2802  
QY 470 TyrSerProIleTyIleArgAspArgSerPheProSerGlnArgAspAspGlnAspGlu 489  
Db 2803 CTCTCCCTCAGGCAATTCAGCCACAGCTACCCAGGA----- 2838  
QY 490 IleSerAlaLeuValSerSerTyrLeuGlyProSerThrSerPheProHisArgSerArg 509  
Db 2839 -----GCTTCAGCTGGTCTCACAAT----- 2859  
QY 510 ArgSerProGluTyrMetValProLeuProHisGlyGlyLeuGlyArgSerValTyrAla 529  
Db 2860 -----CCTACTGGGGGTTTCAGGA----- 2877  
QY 530 TyrGluHisLeuAlaProAsnSerTyrSerProGlyHisGlyHisArgLeuHisArgGln 549  
Db 2878 -----CCTCACTGGGGTTCATCCCATG----- 2898  
QY 550 TyrSerProSerLeuValHisGlyGlnArgHisArgProLeuGlnTyrSerPro----- 566  
Db 2899 -----GGTGGCACACCAAGCTTGGGGTTCATCCCATGAGCGGCGGACCCCAACCATGG 2952  
QY 567 -----ProIleHisGlyGlnGlnLeuProTyrGlyIleGlnArgValTyrArgHis 584  
Db 2953 GGTCAACCCCTCAGGGCCAATGCAAGGGGTACCTCGAGGTAGCAGTATAGGAGTCCGCAAT 3012  
QY 595 SerProSerGluGluArgTyrLeu-----GlyLeuSerAsnGlnArgSerProArg 601  
Db 3013 AGCCCCCAGGCTCTGAGGCGGACAGCTTCTGGGGGAGCGGACGGAACAGGGCATGACAGA 3072  
QY 602 SerAsnSer 604  
Db 3073 AGCAGAGT 3081

RESULT 11

AAZ40489 standard; DNA; 3824 BP.

ID

XX



QY 455 ProArgLeuSerProMetGluMetProValThrSerSerSerTyrSerProIleTyr 474  
 Db 1201 -----CCAAATGCCACCAAGAGCGGCGTATCACA----- 1233  
 QY 475 ArgAspArgSerPheProSerGlnArgAspAspGlnAspGluIleSerAlaLeuVal 494  
 Db 1234 -----AACGGTTACGTC 1245  
 QY 495 SerSerTyrLeuGlyProSerThrSerPheProHisArgSerArgSerProGluTyr 514  
 Db 1246 TCCTCTTCCACCGCTCCACCAACATTCATCAGATCCCAAAATCACTCACCTCAATAC 1305  
 QY 515 MetValProLeuProHisGlyGlyLeuArgSerValTyrAlaTyrGluHisLeuAla 534  
 Db 1306 GGTGTACCA-----GCATACACCATCCCA 1332  
 QY 535 ProAnSerTyr-----SerProGlyHisGlyHisArgLeuHisArgGlnTyrSer 551  
 Db 1333 CTTACCATCTACAGCAACAGAGCGCTCCATAC-----CAATACTCA 1374  
 QY 552 ProSerLeuValHisGly-----GlnArgHisProLeuGlnTyrSerProProIleHis 569  
 Db 1375 CCGAGGCGGTTCATGGTCTTACCAACCTCTCCGTCAGTTAT-----CCGACAGCATAT 1431  
 QY 570 GlyGlnGlnGlnLeuProTyrGlyLe-----GlnArgValTyrArgHisSerProSer 587  
 Db 1432 GGTACTACTGAGTCGCGTGGTCTCCACCCCTCCAGTTTACCATCTCCACCGCAC 1491  
 QY 588 GluGluArgTyrLeu 592  
 Db 1492 CACCACCATAT 1506  
 RESULT 9  
 ID ABL17653  
 XX ABL17653 standard; DNA; 7785 BP.  
 AC ABL17653;  
 XX 26-MAR-2002 (first entry)  
 DT Drosophila melanogaster genomic polynucleotide SEQ ID NO 4432.  
 DE Drosophila; developmental biology; cell signalling; insecticide;  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ds.  
 XX Drosophila melanogaster.  
 OS WO200171042-A2.  
 PN 27-SEP-2001.  
 PD 23-MAR-2001; 2001WO-US09231.  
 PF 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX (PEKE ) PE CORP NY.  
 PA Venter JC, Adams M, Li PWD, Myers EW;  
 PI WPI; 2001-656860/75.  
 DR New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX Claim 1; SEQ ID NO 4432; 21pp + Sequence Listing; English.  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 7785 BP; 2166 A; 2186 C; 2007 G; 1426 T; 0 other;  
 SQ Alignment Scores:  
 Pred. No.: 0.0687 Length: 7785  
 Score: 152.00 Matches: 129  
 Percent Similarity: 37.78% Conservative: 106  
 Best Local Similarity: 20.74% Mismatches: 238  
 Query Match: 4.90% Indels: 149  
 Db: 23 Gaps: 31  
 US-09-890-475-1 (1-609) x ABL17653 (1-7785)  
 QY 5 ProProThrValAlaAlaGlnProThr-----ThrThrAlaAsnProLeuLeuGlnArgHis 23  
 Db 3487 CCTCGTGGTGGCGCAAGTCCAGCGCCGCCAAGCCAATCTTTC----- 3534  
 QY 24 GlnSerGluGlnArgArgGluLeuProLysIleValGluThrGluSerThr----- 41  
 Db 3535 -----GACCAACCCCGCACCCGAGAGCCCGTCAATCCACCTATGCCACCCAGT 3588  
 QY 42 -----SerMetAspIleThrIleGlyGlnSerLys 51  
 Db 3589 CCTATCAGATGGAGACAGCACCCCTGAGACCTGAGATCTCATTGGCAACATGAA 3648  
 QY 52 GlnProGlnPheLeuLysSerIleAspGluLeuAlaAlaPheSerValAlaValGluThr 71  
 Db 3649 CAGAACCGCCCAAGCAGAACCCAGGAGTTCTACGTTCATCGAGACCATGTGGACACA 3708  
 QY 72 PheLysArgGlnPheAspAspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAsp 91  
 Db 3709 TTCAGGGAGAG-----CCGNACTGGAGTACGAAACCGAT 3744  
 QY 92 SerLysLeuGluSerAsnGlyValValLeuAlaAlaArgAsnAsnAsnPheHisGlnPro 111  
 Db 3745 CCGGAGATATATCGT-----ATCCTAACAAAGAGGCCCATTAAGCAATGAACCA 3795  
 QY 112 MetLeuSerProProArgAsnAsnValSerValGluThrThrValThr-----ValSerGln 130  
 Db 3796 -----TCTGACCGCGCAGAGGATAGTCAACCGATGATGTAAGAACTTTGTATGGCAA 3849  
 QY 131 ProSerGlnGluIleValPro-----GluThrSerAsnLys 142  
 Db 3850 CACATTGAGAAATTTGTGCCAACCTGTCACAGGAGGCTCATGGAGAAA 3909  
 QY 143 ProGluGlyGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyrIleTyr 162  
 Db 3910 CCGAAGAGGTAGCTAAATTGAG----- 3933  
 QY 163 AlaAsnIleSerAspGlnAlaLys-----LeuMetGluGluIleProSerAla 178  
 Db 3934 -----GCAGAACAGCCAAAGGAGCCTGAACCTTACTCGAACCTCAGCCTGAGCCC 3984  
 QY 179 LeuLysLeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPheTyrLeu 198  
 Db 3985 ACACCA---AGTGTGATCTCTGTCTACATCATGGAGGAGTCAAGCCCAATTA 4041  
 QY 199 GlnGlyArgAlaPheThrLysGluSerProMetSerSer-----AlaArgGln 215  
 Db 4042 -----AGGGAGCGCAGCATCTCGAGGAAGTTCAACCCAGTTTATCTATGCCACTGAA 4095  
 QY 216 ValSerLeuLeuIleLeuGluSerPheLeuLeuMetProAspArgGlyLysVal 235  
 Db 4096 ATAGCC-----AACTCAAGTTGGAATTTGATCGTGC----- 4128  
 QY 236 LysIleGluSerTrpIleLysAspGluAlaGluThrAlaAlaValAlaTyrArgLysArg 255



## RESULT. T. 8

PR	03-JUN-1999;	99US-0137528.	PR	05-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	99US-0137502.	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	99US-0137724.	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	99US-0138094.	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138540.	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	99US-0138847.	PR	10-AUG-1999;	99US-0148171.
PR	14-JUN-1999;	99US-0139119.	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	99US-0139452.	PR	12-AUG-1999;	99US-0148341.
PR	16-JUN-1999;	99US-0139453.	PR	13-AUG-1999;	99US-0148565.
PR	17-JUN-1999;	99US-0139492.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139454.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139455.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139456.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139457.	PR	18-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139462.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151066.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151080.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	99US-0139899.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	99US-0140354.	PR	01-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	99US-0140635.	PR	07-SEP-1999;	99US-0152363.
PR	28-JUN-1999;	99US-0140823.	PR	10-SEP-1999;	99US-0153070.
PR	29-JUN-1999;	99US-0140931.	PR	13-SEP-1999;	99US-0153758.
PR	30-JUN-1999;	99US-0141287.	PR	15-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	99US-0141842.	PR	16-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	99US-0142154.	PR	20-SEP-1999;	99US-0154779.
PR	02-JUL-1999;	99US-0142055.	PR	22-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142390.	PR	23-SEP-1999;	99US-0155486.
PR	08-JUL-1999;	99US-0142803.	PR	24-SEP-1999;	99US-0155659.
PR	09-JUL-1999;	99US-0142920.	PR	28-SEP-1999;	99US-0156458.
PR	12-JUL-1999;	99US-0142977.	PR	29-SEP-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143542.	PR	04-OCT-1999;	99US-0157117.
PR	14-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144325.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144332.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159337.
PR	20-JUL-1999;	99US-0144884.	PR	14-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144814.	PR	14-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145086.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160989.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161359.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146388.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161920.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161992.
PR	04-AUG-1999;	99US-0147204.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147302.	PR	29-OCT-1999;	99US-0162142.
PR	05-AUG-1999;	99US-0147192.			

Alignment Scores:

QY 162 TyrAlaAsnIleSerAspGluAlaLysLeuMetGluGluIleProSerAlaLeuLysLeu 181  
DB 545 ACGCGGAGAGAGAGAGAGAGATTCGGTCAAGAGATTCGGCGGCTTGGTTGAT 604  
QY 182 AlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPheTyrLeuGlnGlyArg 201  
DB 605 TGTGTAGATCCGGCGATGTTAGTGTGTAAGCGATATCTGAGGTTTTTCCGCTGGATACA 664  
QY 202 ArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuLeu 221  
DB 665 AGAGAGATAAAGTTAGCAATGATATGATGAGGCTTGTGTGTGATCTTGAGAGTTTA 724  
QY 222 GluSerPheLeuLeuMetProAspArgGlyLysGlyLysValLysIleGluSerTrpIle 241  
DB 725 ACACCTGTAAATGTTGATCCATGATGTTGGAAATCGAGGTGCTTGTACTCCGAGTGT 784  
QY 242 LysAspGluAlaGluThrAlaAlaValAlaIleTrrpGlyLysArgLeuMetThrGluGlyGly 261  
DB 785 AAGGAGAAAGCTTAAGGAGATTCCTGAGAGCTGCGAAGAGATTTTGGAGAGAGAGGAGG 844  
QY 262 LeuAlaAlaAlaGluLysMetAspAlaAlaArgGlyLeuLeuLeuValAlaCysPheGly 281  
DB 845 ATAGAGATGTGAACATCTCTGATGTTTATACGTTTTTGCAGCATCTGTATACATTGGA 904  
QY 282 ValProSerAsnPheArgSerThrAspLeuLeuAspLeuLeuIleArgMetSerGlySerAsn 301  
DB 905 ATT-----GTTAAAGTCAAGATCTCTGTTGTATAGAAAGCTT----- 943  
QY 302 GluIleAlaGlyAlaLeuLysArgSerGlnPhe-----Leu 313  
DB 944 ---GTTGTGATCTGCGTGGGTTAAACAGATGCTTAAGCTTGGCTTCTGTGTTGTTG 1000  
QY 314 ValProMetValSerGlyIleValGluSerSerIleLysArgGlyMetHisIleGluAla 333  
DB 1001 GGTGATCAATGCTGACATGATTAAGAAATTAATCAGCAGGAGACACAGCTTGATGCT 1060  
QY 334 LeuGluMetValThrPheGlyMetGluAspLysPheSerAlaAlaLeuValLeuThr 353  
DB 1061 GTTCATTTTACTTATGAATGTTGGCTTGTGATAAGTTCCCTCTGTTCTCTGCTCAAA 1120  
QY 354 SerPheLeuLysMetSerLysGluSerPheGluArg---AlaLysArgLysAlaGlnSer 372  
DB 1121 GCTTATCAGGAGACGCAAGAAAGTCAGCAGCTTCTATCATGAGGAGCTCTGCAATACT 1180  
QY 373 ProLeuAlaPheLysGluAlaAlaThrLysGlnLeuAlaValLeuSerSerValMetGln 392  
DB 1181 GGCAGGCTACGATCTCTGTCGCGCGCAAGAGCAATCAGCAGCTTAAGCGGTTTAAAA 1240  
QY 393 CysMetGluThrHisLysLeuAspProAlaLysGluLeuProGlyTrrpGlnIleLysGlu 412  
DB 1241 TGCATAGAGAGTACAACTCGAG-----GAAGAGTTCCCTCCAGAGAACCTCAAGAA 1294  
QY 413 GlnIleValSerLeuGluLysAspThrLeuGlnLeuAspLysGluMetGluGluLysAla 432  
DB 1295 CGGTTG-----GATCAGCTAGAGAA-GACCAAAACCGAGAGAGAGAACGCC 1338  
QY 433 ArgSerLeuSerLeuMetGluGluAlaLeuAlaLysArgMetTyrAsnGlnIle 452  
DB 1339 AGCAGCTGTCTCTGCCAAAGAAAGACCG-----AGCCAGCTA 1377  
QY 453 LysArg-ProArgLeuSerProMetGluMetProValThrSerSerSerTyrSerPr 472  
DB 1378 CAACGGTCCA-----ATGCCACCGCAAGCGCGGCTATCACAAA 1419  
QY 472 IleTyrArgAspArgSerPheProSerGlnArgAspAspGlnAspGluIleSerAl 492  
DB 1420 CGCATACGTCCTT---TCCTTCCGCTTCAATCAGATCACCTCTCACTCTCTCAATACGC 1476  
QY 492 AleuValSerSerTyrLeuGlyProSerThrSerPheProHisArgSerArgArgSerPr 512  
DB 1477 TTCACAGAGCTTACCATCCCACTTACACAGTC-----TACAGCAACAGAGGCC 1530  
QY 512 oGluTyr-----MetValProLeuProHisGlyGlyLeuGlyArgSerVa 527

DB 1531 ACCTTATCCATCTCTCCGAGATCATCCAGGCTCATACCAAGGCTCTCTCTATCGTTA 1590  
QY 527 lTyrAlaTyrGluHisLeu-----AlaProAsnSerTyrSe 539  
DB 1591 CCCAGCTACAACGGTTATTGTATATGTCAGTTCCTGCTCCAGCTCTCTCCAGTTTACCA 1650  
QY 539 rProGlyHisGlyHisArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGlnAr 559  
DB 1651 CCCA---CACCATCATCAA---CACCACCAATTC-----CACCATCAGCA 1689  
QY 559 gHis 560  
DB 1690 ACAT 1693  
RESULT 7  
AAC38630  
ID AAC38630 standard; DNA; 1865 BP.  
XX AAC38630;  
XX 17-OCT-2000 (first entry)  
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 21672.  
XX Hybridisation assay; genetic mapping; gene expression control;  
XX protein identification; signal transduction pathway;  
XX metabolic pathway; promoter; termination sequence; ss.  
XX Arabidopsis thaliana.  
XX EF1033405-A2.  
XX 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-0301439.  
XX 25-FEB-1999; 99US-0121825.  
XX 05-MAR-1999; 99US-0123180.  
XX 09-MAR-1999; 99US-0123548.  
XX 23-MAR-1999; 99US-0125788.  
XX 25-MAR-1999; 99US-0126264.  
XX 29-MAR-1999; 99US-0126785.  
XX 01-APR-1999; 99US-0127462.  
XX 06-APR-1999; 99US-0128234.  
XX 08-APR-1999; 99US-0128714.  
XX 16-APR-1999; 99US-0129845.  
XX 19-APR-1999; 99US-0130077.  
XX 21-APR-1999; 99US-0130510.  
XX 23-APR-1999; 99US-0130891.  
XX 28-APR-1999; 99US-0131449.  
XX 30-APR-1999; 99US-0132048.  
XX 30-APR-1999; 99US-0132407.  
XX 04-MAY-1999; 99US-0132484.  
XX 05-MAY-1999; 99US-0132485.  
XX 06-MAY-1999; 99US-0132486.  
XX 07-MAY-1999; 99US-0132487.  
XX 07-MAY-1999; 99US-0132863.  
XX 11-MAY-1999; 99US-0134256.  
XX 14-MAY-1999; 99US-0134218.  
XX 14-MAY-1999; 99US-0134219.  
XX 14-MAY-1999; 99US-0134221.  
XX 14-MAY-1999; 99US-0134370.  
XX 18-MAY-1999; 99US-0134768.  
XX 19-MAY-1999; 99US-0134941.  
XX 20-MAY-1999; 99US-0135124.  
XX 21-MAY-1999; 99US-0135353.  
XX 24-MAY-1999; 99US-0135629.  
XX 25-MAY-1999; 99US-0136021.  
XX 27-MAY-1999; 99US-0136392.  
XX 28-MAY-1999; 99US-0136782.  
XX 01-JUN-1999; 99US-0137222.

PR	04-OCT-1999;	99US-0157117.	
PR	05-OCT-1999;	99US-0157753.	
PR	06-OCT-1999;	99US-0157865.	
PR	07-OCT-1999;	99US-0158029.	
PR	08-OCT-1999;	99US-0158232.	
PR	12-OCT-1999;	99US-0158369.	
PR	13-OCT-1999;	99US-0159293.	
PR	13-OCT-1999;	99US-0159294.	
PR	13-OCT-1999;	99US-0159295.	
PR	14-OCT-1999;	99US-0159329.	
PR	14-OCT-1999;	99US-0159330.	
PR	14-OCT-1999;	99US-0159331.	
PR	14-OCT-1999;	99US-0159637.	
PR	14-OCT-1999;	99US-0159638.	
PR	18-OCT-1999;	99US-0159584.	
PR	21-OCT-1999;	99US-0160741.	
PR	21-OCT-1999;	99US-0160767.	
PR	21-OCT-1999;	99US-0160768.	
PR	21-OCT-1999;	99US-0160770.	
PR	21-OCT-1999;	99US-0160814.	
PR	21-OCT-1999;	99US-0160815.	
PR	22-OCT-1999;	99US-0160980.	
PR	22-OCT-1999;	99US-0160981.	
PR	22-OCT-1999;	99US-0160989.	
PR	25-OCT-1999;	99US-0161404.	
PR	25-OCT-1999;	99US-0161405.	
PR	25-OCT-1999;	99US-0161406.	
PR	26-OCT-1999;	99US-0161359.	
PR	26-OCT-1999;	99US-0161360.	
PR	26-OCT-1999;	99US-0161361.	
PR	28-OCT-1999;	99US-0161920.	
PR	28-OCT-1999;	99US-0161992.	
PR	28-OCT-1999;	99US-0161993.	
PR	29-OCT-1999;	99US-0162142.	
Alignment Scores:			
Pred. No.:	1.55e-14	Length:	1970
Score:	296.50	Matches:	130
Percent Similarity:	41.10%	Conservative:	101
Best Local Similarity:	23.13%	Mismatches:	231
Query Match:	9.55%	Indels:	101
DB:	21	Gaps:	19
US-09-890-475-1 (1-609) x AAC35353 (1-1970)			
Qy	59	IleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGln-----	75
Db	131	TTAACTGAGTTAGCT-----CAGCCGAGTTTCGAGGAGTTTCAGAAACAGACGTCGTTG	184
Qy	76	-----PheAspAspLeuGlnLysHisIleGluSerIleGluAsn	88
Db	185	ATGACTAGCTGTACTCTTCTATGCGCAAGAGCTCTCCGATCACTTCCTCTAGAGCAG	244
Qy	89	AlaIleAspSerLysLeuGluSerAsnGlyValValLeuAlaAlaArgAsnAsnPhe	108
Db	245	AATCTGATGAAGAAATCTGAAGCGTTGAACGAGATGATTGAACCTAGATAACCGAGCT	304
Qy	109	HisGlnProMetLeuSerProProArgAsnAsnValSerValGluThrThrValThrVal	128
Db	305	CAGACCTCGCTCGAGTCTCTGAAGCGTAGAGAGGTAAACATCGACCATAGCGTTGAGATC	364
Qy	129	-----SerGlnProSerGln	133
Db	365	GTAGCTGGGAAAGTAGGTGAACGAGCTAGAGTGTCTCTTGAATCGTTAGAGAAAGCTAGA	424
Qy	134	GluIleValProGluThrSerAsnLysProGluGlyGly-----	146
Db	425	GATGGTTGTGTGATGGTTCTTAATGATGATTCGCGTGATGTTGACGATGAGAGAGGTCTT	484
Qy	147	-----ArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyrIle	161
Db	485	CTCTCTGCTTTGAAATCTCTTTGCTTAAATGGATGCGAGAGGTTTTTGGAACTTTGTG	544

PR	25-FEB-1999	99US-01218295
PR	05-MAR-1999	99US-01231801
PR	09-MAR-1999	99US-01235488
PR	23-MAR-1999	99US-01257088
PR	25-MAR-1999	99US-01262624
PR	29-MAR-1999	99US-01267895
PR	01-APR-1999	99US-01274652
PR	01-APR-1999	99US-01274652
PR	06-APR-1999	99US-01282834
PR	08-APR-1999	99US-01287845
PR	16-APR-1999	99US-01298077
PR	19-APR-1999	99US-01300077
PR	21-APR-1999	99US-01303449
PR	23-APR-1999	99US-01305018
PR	23-APR-1999	99US-01308911
PR	28-APR-1999	99US-01311449
PR	30-APR-1999	99US-01320048
PR	30-APR-1999	99US-01324047
PR	04-MAY-1999	99US-01324484
PR	05-MAY-1999	99US-01324485
PR	06-MAY-1999	99US-01324486
PR	06-MAY-1999	99US-01324487
PR	11-MAY-1999	99US-01328863
PR	11-MAY-1999	99US-01342186
PR	14-MAY-1999	99US-01342186
PR	14-MAY-1999	99US-01342211
PR	14-MAY-1999	99US-01343708
PR	18-MAY-1999	99US-01343708
PR	19-MAY-1999	99US-01349491
PR	20-MAY-1999	99US-01351241
PR	21-MAY-1999	99US-01353553
PR	24-MAY-1999	99US-01356629
PR	25-MAY-1999	99US-01360621
PR	27-MAY-1999	99US-01363932
PR	28-MAY-1999	99US-01367682
PR	01-JUN-1999	99US-01372722
PR	03-JUN-1999	99US-01375288
PR	04-JUN-1999	99US-01375002
PR	07-JUN-1999	99US-01377724
PR	07-JUN-1999	99US-01380094
PR	08-JUN-1999	99US-01385409
PR	10-JUN-1999	99US-01388847
PR	10-JUN-1999	99US-01391119
PR	14-JUN-1999	99US-01394577
PR	16-JUN-1999	99US-01394577
PR	16-JUN-1999	99US-01394583
PR	17-JUN-1999	99US-01394992
PR	17-JUN-1999	99US-01394954
PR	18-JUN-1999	99US-01394555
PR	18-JUN-1999	99US-01394556
PR	18-JUN-1999	99US-01394560
PR	18-JUN-1999	99US-01394611
PR	18-JUN-1999	99US-01394621
PR	18-JUN-1999	99US-01394631
PR	18-JUN-1999	99US-01397500
PR	18-JUN-1999	99US-01394588
PR	18-JUN-1999	99US-01394599
PR	21-JUN-1999	99US-01398177
PR	22-JUN-1999	99US-01398699
PR	23-JUN-1999	99US-01403533
PR	23-JUN-1999	99US-01403554
PR	24-JUN-1999	99US-01406955
PR	24-JUN-1999	99US-01420555
PR	02-JUL-1999	99US-01423900
PR	06-JUL-1999	99US-01428300
PR	09-JUL-1999	99US-01439320
PR	12-JUL-1999	99US-01439377

PR	06-AUG-1999;	99US-0147303.	Score:	296.50	Matches:	130
PR	06-AUG-1999;	99US-0147416.	Percent Similarity:	41.10%	Conservative:	101
PR	09-AUG-1999;	99US-0147493.	Best Local Similarity:	23.13%	Mismatches:	231
PR	09-AUG-1999;	99US-0147935.	Query Match:	9.55%	Indels:	101
PR	10-AUG-1999;	99US-0148171.	DB:	21	Gaps:	19
PR	11-AUG-1999;	99US-0148319.	US-09-890-475-1 (1-609) x AAC49876 (1-1964)			
PR	12-AUG-1999;	99US-0148341.	Qy	59	IleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLysA:ggIn-----	75
PR	13-AUG-1999;	99US-0148565.	Db	128	TTAACTGAGTTAGCT-----CAGCCGAGTTTCGAGGAGTTTCAGAAACACAGACGTCGGTTG	181
PR	16-AUG-1999;	99US-0149368.	Qy	76	-----PheAspAspLeuGlnLysHisIleGluSerIleGluAsn	88
PR	18-AUG-1999;	99US-0149426.	Db	182	ATGACTAGCTGTAAGCTTCTATGCGAAGAGCTCCGATCCTCCTCCTCTCTAGAGCAG	241
PR	20-AUG-1999;	99US-0149723.	Qy	89	AlaIleAspSerLysLeuGluSerAsnGlyValValLeuAlaAlaArgAsnAsnAsnPhe	108
PR	20-AUG-1999;	99US-0149729.	Db	242	AATCTGATGAAGAAATCTGAAGCGTTGAAGCAGATGATTGAACCTCTAGATAACCAACT	301
PR	23-AUG-1999;	99US-0149902.	Qy	109	HisGlnProMetLeuSerProProArgAsnValSerValGluThrValThrVal	128
PR	23-AUG-1999;	99US-0149930.	Db	302	CAGACCTCGCTCGAGTCTCTGAAGCGTAGAGAGGTAAACATCGACCATAGCGTTGAGATC	361
PR	25-AUG-1999;	99US-0150566.	Qy	129	-----SerGlnProSerGln	133
PR	26-AUG-1999;	99US-0150884.	Db	362	GTAGCTGGGAAGTAGTGAACGAGTAGAGCTGCTCTTGTGATCGTTAGAGAAAGCTAGA	421
PR	27-AUG-1999;	99US-0151065.	Qy	134	GluIleValProGluThrSerAsnLysProGluGlyGly-	146
PR	27-AUG-1999;	99US-0151066.	Db	422	GATGTTGTGTGATGCTTAATGATGATTCGGTGATGTTGACGATGAAGAAGGTCTT	481
PR	27-AUG-1999;	99US-0151080.	Qy	147	-----ArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTrpIle	161
PR	30-AUG-1999;	99US-0151303.	Db	482	CTTTCTGCTTTGAAATCTCTTTGCTTAAATGGAATCGAGAGGTTTTTGGAACTTTGTG	541
PR	31-AUG-1999;	99US-0151438.	Qy	162	TyrAlaAsnIleSerAspGlnAlaLysLeuMetGluIleProSerAlaLeuLysLeu	181
PR	01-SEP-1999;	99US-0151930.	Db	542	ACGCGAGGAGAGAGGAGTGGAGAAATCTCCGGTCAAAGATTCCGGCGCGTGTGTTGAT	601
PR	07-SEP-1999;	99US-0152363.	Qy	182	AlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPheTyLeuGlnGlyArg	201
PR	10-SEP-1999;	99US-0153070.	Db	602	TGTGTAGATCCGCGGATGTTAGTGTCTTGAAGCGATATCTAGGTTTTTCCCGTGGATACA	661
PR	13-SEP-1999;	99US-0153758.	Qy	202	ArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuLeu	221
PR	15-SEP-1999;	99US-0154018.	Db	662	AGAGGAGATAAAGTTAGCAATGATTATGATCGGCTTGTGTTGTGATCTTCTGAGAGTTA	721
PR	16-SEP-1999;	99US-0154039.	Qy	222	GluSerPheLeuLeuMetProAspArgGlyLysGlyLysValLysValLysIleGluSerTrpIle	241
PR	22-SEP-1999;	99US-0154779.	Db	722	ACACCTGTAATTGTTGATCCAGTGAATGGGAAATCGAGGTTGCTTTACTCCGAGTGT	781
PR	23-SEP-1999;	99US-0155139.	Qy	242	LysAspGluAlaGluThrAlaAlaValAlaAlaTrpArgLysArgLeuMetThrGluGlyGly	261
PR	23-SEP-1999;	99US-0155486.	Db	782	AGGAGAAAGCTAAGAGATGCTGAGACGTGAAGACAGAGTTTGAAGAGAGAGAGG	841
PR	24-SEP-1999;	99US-0156559.	Qy	262	LeuAlaAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuValAlaCysPheGly	281
PR	28-SEP-1999;	99US-0156596.	Db	842	ATAGAGAATCTGAAGACTCCTGATGTTTCATACGTTTTTTCAGCATCTTTGTACATTGGA	901
PR	04-OCT-1999;	99US-0157117.	Qy	282	ValProSerAsnPheArgSerThrAspLeuLeuAspLeuIleArgMetSerGlySerAsn	301
PR	06-OCT-1999;	99US-0157865.	Db	902	ATT-----GTTAAAGTGAAGATCTTGTGTTGTATAGAAAGCTT-----	940
PR	07-OCT-1999;	99US-0158029.	Qy	302	GluIleAlaGlyAlaLeuLysArgSerGlnPhe-----Leu	313
PR	08-OCT-1999;	99US-0158232.	Db	941	---GTTCTTGGATCTCGGTGGCGTAAACACAGATGCCCTTAAGCTTGGCGTTTCTCTCTGTTG	997
PR	12-OCT-1999;	99US-0158369.	Qy	314	ValProMetValSerGlyIleValGluSerSerIleLysArgGlyMetHisIleGluAla	333
PR	13-OCT-1999;	99US-0159293.	Db	998	GGTGATCAATGCTGACATGATTGAAGAAATTAATCAGCAGGGGCAACAGCGTTGATGCT	1057
PR	13-OCT-1999;	99US-0159294.	Qy	334	LeuGluMetValTyrThrPheGlyMetGluAspLysPheSerAlaAlaLeuValLeuThr	353
PR	14-OCT-1999;	99US-0159330.	Db			
PR	14-OCT-1999;	99US-0159637.	Qy			
PR	14-OCT-1999;	99US-0159638.	Db			
PR	18-OCT-1999;	99US-0159584.	Qy			
PR	21-OCT-1999;	99US-0160741.	Db			
PR	21-OCT-1999;	99US-0160767.	Qy			
PR	21-OCT-1999;	99US-0160768.	Db			
PR	21-OCT-1999;	99US-0160770.	Qy			
PR	21-OCT-1999;	99US-0160814.	Db			
PR	21-OCT-1999;	99US-0160815.	Qy			
PR	22-OCT-1999;	99US-0160980.	Db			
PR	22-OCT-1999;	99US-0160981.	Qy			
PR	22-OCT-1999;	99US-0160989.	Db			
PR	25-OCT-1999;	99US-0161404.	Qy			
PR	25-OCT-1999;	99US-0161405.	Db			
PR	25-OCT-1999;	99US-0161406.	Qy			
PR	26-OCT-1999;	99US-0161359.	Db			
PR	26-OCT-1999;	99US-0161360.	Qy			
PR	26-OCT-1999;	99US-0161361.	Db			
PR	28-OCT-1999;	99US-0161920.	Qy			
PR	28-OCT-1999;	99US-0161992.	Db			
PR	28-OCT-1999;	99US-0161993.	Qy			
PR	29-OCT-1999;	99US-0162142.	Db			

Alignment Scores:  
Pred. No.: 1.55e-14 Length: 1964

Db 1748 AGTCATCCCTGAAGG---TATCCGAGTATGCTATGACACACAGACCGTTCCTTAGCGG 1804  
Qy 492 aLeuValSerTyrLeuGlyProSerThrSerPheProHisArgSerArgArgSerPi 512  
Db 1805 TCCAATCGGCGACACACACCTCCACCACTCTCTCTCTCAGACTTACACGTTCAATCC 1864  
Qy 512 oGluTyrMetValProLeuProHisGlyGlyLeuGlyArgSerValTyrAlaTyrGluHi 532  
Db 1865 T-----GTCCT---GCATCGGGAAGTCTTACCGCACTGCTATCAGTACCGGC 1912  
Qy 532 sLeuAlaProAnSerTyr 538  
Db 1913 TCCTCTCTCTCTCTCTTAC 1931  
RESULT 5  
AAC49876  
ID AAC49876 standard; DNA; 1964 BP.  
XX AAC49876;  
XX  
XX 18-OCT-2000 (first entry)  
DT Arabidopsis thaliana DNA fragment SEQ ID NO: 62772.  
DE  
XX  
XX Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss  
XX  
XX Arabidopsis thaliana.  
XX  
XX EP1033405-A2.  
XX  
XX  
XX 06-SEP-2000.  
XX  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 03-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139484.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.

PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0158393.  
 PR 13-OCT-1999; 99US-0158294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160880.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

## Alignment Scores:

Pred. No.: 4,848-18 Length: 2158  
 Score: 340.00 Matches: 139  
 Percent Similarity: 41.13% Conservative: 86  
 Best Local Similarity: 25.41% Mismatches: 217  
 Query Match: 10.95% Indels: 105  
 DB: 21 Gaps: 21

US-09-890-475-1 (1-609) x AAC46135 (1-2158)

QY 64 AlapheSerValAlaValGluThrPhe-LysArgGlnPhe-----AspAspLeuG1 80  
 DB 392 GCATTTTCATGGCTTGAGAGGTTCTTTGAAGAGAGGTTTCATGAGCTTGAAGATCAAGA 451  
 QY 80 nlyshHistleGluSerlleGluAsnAlaIleAspSerLysLeuGluSerAsnGlyValVa 100  
 DB 452 AAAGAGATGAACCAAAACAGAAAGCTCAAGAACTGTTGGAGAAAAGAGGAGC 511  
 QY 100 lleuAlaIalaArgAsn----- 105  
 DB 512 TGTTCAGGCTAAGGAGAGCGCGCTTGAAGGCTTCAGAGAGAGAGAGACCGAGTAT 571  
 QY 106 -----AsnAsnPheHisGlnProMetLeuSerPr 115  
 DB 572 GTTTACTATAACAGTGCTTTGGATAAGTATAACAC-----GCTCCTGTCAGTAAACC 625  
 QY 115 O-----ProArgAsn-----AsnValSerValG1 123  
 DB 636 TTCTGTTGGTGAACGTTGGCCACAAATGCTGTAGAGGATTCACATGCTTTGCTGC 685  
 QY 123 uThrThrValThrValSerGlnPro-----SerGlnGluValPro-- 137  
 DB 686 TGATGATTATACCGATGATAATCCGGATGGTATTGTGCAAGATGTTTCAGATCTCACCTGT 745  
 QY 138 -----GluThrSerAsnLysProGluGlyGlyArgMetCysGluLeuMetCy 153

DB 746 CATGGAAATATAGGTGAAGGCTTATCCACAGTTACTGAACACTGTGCGGGATATGA 805  
 QY 153 sserLysGlyLeuArgLysTyrIleTyAlaAsnIleSerAspGlnAlaLysLeuMetG1 173  
 DB 806 CTCACGGGTCTCCATAAGTTTGTATCAGATAACCCGCAAGAACCTTGCATCACTAAGA 865  
 QY 173 uGluIleProSerAlaLeuLysLeuAlaLysGluProAlaLysPheValLeuAspCysI1 193  
 DB 866 GGAGATTCCTATGGCGTTCAGGCTGCAACACCGAGCAGCTTAGTGTGACTCTTT 925  
 QY 193 eGlyLysPheTyr-----LeuGlnGlyArgArgAlaPheThrLysG1 207  
 DB 926 GGAAGGGTTTACCTATGGAGGCACCACTGCTGATGGAATA-----AAAGA 973  
 QY 207 userProMetSerSerAlaArgGlnValSerLeuLeuLeu-----SerPheLe 225  
 DB 974 CGCTAACCTCTTGGGCATGCGCAGAACCTGTATCATGTGTGATGGAGTGCTTAGCATACT 1033  
 QY 225 uLeuMetPro---AspArgGlyLysGlyLysValLysIleGluSerTrpIleLysAspG1 244  
 DB 1034 GTTATCTGGTCTGGACCGCAACTGCTGTGTCTCTCACAATAATGTTAAGCATAG 1093  
 QY 244 uAlaGluThrAlaAlaValAlaIleArgLysArgLeuMetThr---GluGlyGlyLeuAl 263  
 DB 1094 AGCCAAACATTTGCAGAGAGATGGAATCCACTGCTGGAATCTCTTGACATGGATGCTTG 1153  
 QY 263 aAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuValAlaCysPheGlyValPr 283  
 DB 1154 CAATGGTAATCTTGGAGGCTCATGCATCTCTGCAACTGCTGGCCACTTTTGTCTATTGT 1213  
 QY 283 oSerAsnPheArgSerThrAspLeuLeuAspLeuIleArgMetSerGlySerAsnGluI1 303  
 DB 1214 TGTGATTTTAAAGAAAGATGAACCTCTTAAAGCTGATCCCCATGGTTTCACGTGCGCGTCA 1273  
 QY 303 eAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetValSerGlyLeValGluSe 323  
 DB 1274 GCGAGCTGAGCTCTGCGCTTCACTTGGATTAGTGAATAAATGCCGTGGTGTGATGAAGT 1333  
 QY 323 rSerIleLysArgGlyMethHisIleGluAlaLeuGluMetValTyrThrPheGlyMetG1 343  
 DB 1334 TCTCTGTAACAGCTGGAACAGATGATGCGGTAACTTGGCATTTTGGCTTTTGAACCTCAC 1393  
 QY 343 uAspLysPheSerAlaAlaLeuValLeuThrSerPheLeuLysMetSerLysGluSerPh 363  
 DB 1394 AGAAGCTTCTCACCTGCTCATTAAGTAACTTACTTGTAGAGGCAAGCGATCTTC 1453  
 QY 363 eGluArgAlaLysArgLysAlaGlnSerProLeu-----AlaPheLysGluAlaAlaTh 381  
 DB 1454 CCCCCAAGGCGAGACCTGGCAATCATCTCTGCTGTTCAGGCAAGTATGATGATCAATGA 1513  
 QY 381 rLysGluLeuAlaValLeuSerSerValMetGlnCysMetGluThrHisLysLeuAspPr 401  
 DB 1514 GCGGGAGCTTATAGGCTCAAAACCGTTATAAGTGTATCGAAGAGCATAGGCTGGAA-- 1571  
 QY 401 oAlaLeuGluLeuProGlyTrpGlnIleLysGluIleValSerLeuGluLysAspTh 421  
 DB 1572 ----GACGAGTACCCA-----GTTGAGCCACTTCACAACGGAT 1606  
 QY 421 rLeuGlnLeuAspLysGluMetGluGluLysAlaArgSerLeuSerLeuMetGluGluAl 441  
 DB 1607 TCTCCAGCTCGAAGAAAGCCCAAGCGGACAAAGAGAGAGCAACAGAACCCCATGAAA-- 1661  
 QY 441 aAlaLeuAlaLysArgMetTyrAsnGlnGlnIleLysArgProArgLeuSerProMetG1 461  
 DB 1662 -----CCTCAGCCAAACAGCACCGTGGTGT----- 1688  
 QY 461 uMetProProValThrSerSerSer-----TyrSerPr 472  
 DB 1689 -CAACCCCGAGTCACTGACAAACAAACAAACAACTCAACAAACAAAGACAGGATACGGTAG 1747  
 QY 472 oIleTyrArgAspArgSerPheProSerGlnArgAspAspGlnAspGluIleSerAl 492

XX	25-FEB-1999;	99US-0121825.
PF	25-FEB-1999;	99US-0121825.
XX	05-MAR-1999;	99US-0123180.
PR	05-MAR-1999;	99US-0123180.
PR	09-MAR-1999;	99US-0123548.
PR	23-MAR-1999;	99US-0125788.
PR	23-MAR-1999;	99US-0126264.
PR	29-MAR-1999;	99US-0126785.
PR	01-APR-1999;	99US-0127462.
PR	06-APR-1999;	99US-0128234.
PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138647.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	21-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	24-JUN-1999;	99US-0140823.
PR	28-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142350.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149922.
PR	23-AUG-1999;	99US-0149930.
PR	23-AUG-1999;	99US-0149932.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.

PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

## Alignment Scores:

Pred. No.: 8,86e-19 Length: 2155  
 Score: 349.00 Matches: 138  
 Percent Similarity: 41.10% Conservatives: 86  
 Best Local Similarity: 25.32% Mismatches: 218  
 Query Match: 11.24% Indels: 103  
 DB: 21 Gaps: 20

US-09-890-475-1 (1-609) x AAC33344 (1-2155)

Qy 64 AlaphSerValAlaValGluThrPhe-LysArgGlnPhe-----AspAspLeuGl 80  
 |||||  
 Db 392 GCATTTTCATGGGCTTGAGAGGCTTTGAAGAGAGGTTTCATGAGCTTTGAAGATCAAGA 451  
 |||||  
 Qy 80 nLysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsnGlyValVa 100  
 |||||  
 Db 452 AAAGAGTATGAACCAACCAACAGGAAGCTCAAGAACTGTTGGAGAAAAGAGGCGAGC 511  
 |||||  
 Qy 100 lLeuAlaAlaArgAsn----- 105  
 |||||  
 Db 512 TGTGAGGCTTAAGGAGAGGCGGCTTGAAAGGCTTCAGAGAGAGAGAGAGCGCAGCTAT 571  
 |||||  
 Qy 106 -----AsnAsnPheHisGlnProMetLeuSerPr 115  
 |||||  
 Db 572 GTTTACTATAACAGTGCTTTGGATAGTATATAAAC-----GCTCCTGTGCTAGTAAC 625  
 |||||  
 Qy 115 o-----ProArgAsn-----AsnValSerValGl 123  
 |||||  
 Db 626 TTCTGTTGGTGAACGTTGGCCACAAATGCTGTAGAGGATTCATCCATGCTTTGTGCTG 685  
 |||||  
 Qy 123 uThrThrValThrValSerGlnPro-----SerGlnGluIleValPro-- 137  
 |||||  
 Db 686 TGATAGTATTACCATGATTAATCCGATGTTATGTGCAAGATGTTTCAGATCTCACCTGT 745  
 |||||  
 Qy 138 -----GluThrSerAsnLysProGluGlyGlyArgMetCysGluLeuMetCy 153  
 |||||  
 Db 746 GATGGAAATATTAGGTGAAGGCTTATCCACAGTTACTGAAACTGTGGGGGATATGGA 805  
 |||||  
 Qy 153 sSerLysGlyLeuArgLysThrIleTyAlaAnIleSerAspGlnAlaLysLeuMetGl 173  
 |||||  
 Db 806 CTCAACGGGCTTCATTAAGTTTGTATCAGATTAACCCGCAAGAACCTTGCATCTCAAGGA 865  
 |||||  
 Qy 173 uGluIleProSerAlaLeuLysLeuAlaLysGluProAlaLysPheValLeuAspCysIl 193  
 |||||  
 Db 866 GGAGATTCTATGCGGTTCAGGCTGCGCAACCCAGCCAGCTTAGTGTGGACTCTTT 925  
 |||||  
 Qy 193 eGlyLysPheTy-----LeuGlnGlyArgAlaPheThrLysGl 207  
 |||||  
 Db 926 GGAAGGGTTTTACCCCTATGAGGAGCCCAACTGCTGATGGAANA-----AAAGA 973  
 |||||  
 Qy 207 uSerProMetSerSerAlaArgGlnValSerLeuLeuLeuLeu-----SerPheLe 225  
 |||||  
 Db 974 CGTAACTCTTGGGATGCCAGAACCTGTATCATGTTGATGAGTGGCTTAGCATACT 1033  
 |||||  
 Qy 225 uLeuMetPro---AspArgGlyLysGlyLysValLysIleGluSerTrpIleLysAspGl 244  
 |||||  
 Db 1034 GTTATCTGCTGTCGACCGCAACTCCCTGCTGTGTTCTCTCACAAAATGTTAAGCATAG 1093  
 |||||  
 Qy 244 uAlaGluThrAlaAlaValAlaTrpArgLysArgLeuMetThr--GluGlyGlyLeuAl 263  
 |||||  
 Db 1094 AGCCAAACTATTGCAAGAGATGAATCCACTGCTGGAATCTCTTGAATGATGATGCTTG 1153  
 |||||  
 Qy 263 aAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuLeuValAlaCysPheGlyValPr 283  
 |||||  
 Db 1154 CAATGTTAATCTTTGGAGGCTCATGATCTCTGCAACTGCTGGCCACTTTTCTATTCT 1213  
 |||||  
 Qy 283 oSerAsnPheArgSerThrAspLeuLeuAspLeuLeuLeuLeuLeuSerGlySerAsnGluIl 303  
 |||||

Db 1214 TGCTGATTTTAAAGAGATGAACCTTTAAAGCTGATCCCATGGTTTACGTCGCGTCA 1273  
 Qy 303 eAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetValSerGlyIleValGluSe 323  
 |||||  
 Db 1274 GGCAGCTGAGCTCTGCGCTTCACTTGGATTAGCTGAAAAAATGCTGTGTGATTGAAGT 1333  
 |||||  
 Qy 323 rSerIleLysArgGlyMetHisIleGluAlaLeuGluMetValTyThrPheGlyMetGl 343  
 |||||  
 Db 1334 TCTGTGAACAGTGGAAAAACAGATTGATCGGTTAACTTGGCAATTGCGTTTGAACAC 1393  
 |||||  
 Qy 343 uAspLysPheSerAlaAlaLeuValLeuThrSerPheLeuLysMetSerLysGluSerPh 363  
 |||||  
 Db 1394 AGAACAGTTCTCACCTGTCTCTTACTTAAATCTTACTGTAGAGCAAGCGCATCTTC 1453  
 |||||  
 Qy 363 eGluArgAlaLysArgLysAlaGlnSerProLeuAlaPheLysGluAlaAlaThrLysGl 383  
 |||||  
 Db 1454 CCCCCAAGCAGCAGCTGGCAATGCAATCTCTGCTGTTCCAGATGAGTTCATGAGCGGGA 1513  
 |||||  
 Qy 383 nLeuAlaValLeuSerSerValMetGlnCysMetGluThrHisLysLeuAspProAlaLys 403  
 |||||  
 Db 1514 GCTTATAGGCTTCAAAACCGTTTATAAGTGTATCGAAGAGCATAGCCTGGAA-----GA 1567  
 |||||  
 Qy 403 sGluLeuProGlyTyTrpGlnIleLysGluGlnIleValSerLeuGluLysAspThrLeuGl 423  
 |||||  
 Db 1568 GCAGTACCCA-----GTTAGCCACTTCACAAACGGATTCTCCA 1606  
 |||||  
 Qy 423 nLeuAspLysGluMetGluGluLysAlaArgSerLeuSerLeuMetGluGluAlaAlaLe 443  
 |||||  
 Db 1607 GCTCGAGAAAGCAAGCGGACAAAGAGAGAGCTACAGAACCCATGAAA----- 1655  
 |||||  
 Qy 443 uAlaLysArgMetTyAsnGlnGlnIleLysArgProArgLeuSerProMetGluMetPr 463  
 |||||  
 Db 1656 -----CCTCAGCCAAAACGACACGCTGGTCT-----CAACC 1697  
 |||||  
 Qy 463 oProValThrSerSerSer-----TySerProIleTy 474  
 |||||  
 Db 1688 CCGAGTCACTGACACACACACACACATCAACACAAAGACAGATACGTCAGATCAT 1747  
 |||||  
 Qy 474 rArgAspArgSerPheProSerGlnArgAspAspGlnAspGluIleSerAlaLeuVa 494  
 |||||  
 Db 1748 CCCTGAAAGG---TATCCGACATGTCTATGACACAGACCGTTCCTTAGCGGTCCAAT 1804  
 |||||  
 Qy 494 lSerSerTyLeuGlyProSerThrSerPheProHisArgSerArgArgSerProGluTy 514  
 |||||  
 Db 1805 CATGGCAGCACACCTCCACACCTCTCTCTCTCAGACTTACAGCTTCAATCCT----- 1859  
 |||||  
 Qy 514 rMetValProLeuProHisGlyGlyLeuGlyArgSerValTyAlaTyArgGluHisLeuAl 534  
 |||||  
 Db 1860 ---GCTCCT---GCACATGGGAACTTCTACGCGAACTGCTATCAGTACCAGGCTCCTCC 1912  
 |||||  
 Qy 534 aProAsnSerTyTyr 538  
 |||||  
 Db 1913 TCCTCTCTCTTAC 1925  
 |||||  
 RESULT 4  
 AAC46135  
 ID AAC46135 standard; DNA; 2158 BP.  
 XX AAC46135;  
 AC AAC46135;  
 XX  
 DT 18-OCT-2000 (first entry)  
 XX  
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 49031.  
 KW Hybridisation assay; genetic mapping; gene expression control;  
 KW protein identification; signal transduction pathway;  
 KW metabolic pathway; promoter; termination sequence; ss.  
 XX Arabidopsis thaliana.  
 PV EP1033405-A2.  
 XX  
 PD 06-SEP-2000.

PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 09-JUL-1999; 99US-0142977.  
PR 12-JUL-1999; 99US-0143542.  
PR 13-JUL-1999; 99US-0143624.  
PR 14-JUL-1999; 99US-0144005.  
PR 15-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 20-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145216.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147132.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149358.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160988.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.

Db 1474 AATGAGATTCCGGTCTTTGAAGCGGTACAGTTCTTCTGTCCTATGGTCTCAGGTACC 1533  
Qy 318 ----- 318  
Db 1534 ATATTCGTCTCAGTCGGTGAATTTCAATGCAAGGTGGTTCCTTTGTTGTCATCATC 1593  
Qy 318 ----- 318  
Db 1594 GACCAACATCAAGTCCATCTTTGTTTTCGATAAGCTTGATGGTATATAAATAGGAGAGC 1653  
Qy 318 ----- 318  
Db 1654 ACATCAATATTTAGAGTGCAATGACTGATGAGCCAAATCCTAGCTAGAAATTAATCTG 1713  
Qy 318 ----- 318  
Db 1714 GAAAGAACTTGGAACTCTCAACCATAGTGTGTTGGTACGAAATGTTGCTTGTTCAGAACCA 1773  
Qy 318 ----- 318  
Db 1774 AATGATAGGCTATTGCCTTGAATAGTGTCTTGTGGTTTCCAAATATTGGAAGTTAAAA 1833  
Qy 318 ----- 318  
Db 1834 TCGTATGACTTAGCTGTGGTACTAATTAAGCTTAAGCAATGCCAACTCTAAGAAGTGG 1893  
Qy 319 ----- GlyIleValGluSerSerIleLysArgGlyMet 329  
Db 1894 TACTTACACATATCTTATTTGGTCATAGGTATAGTTGAATCAAGTATCAAGCGTGAATG 1953  
Qy 330 HisIleGluAlaLeuGluMetValTyrPheGlyMetGluAspLysPheSerAlaAla 349  
Db 1954 CATATTGAAGCTCTTGAGTGTGTTTATACCTTTGGCATGGAGATAAGTTTTCAGCTGCT 2013  
Qy 350 LeuValLeuThrSerPheLeuLysMetSerLysGluSerPheGluArgAlaLysArgLys 369  
Db 2014 CTAGTTCTTAACTTCATTTCTTAAAGTGAAGCAGAGTCATTTGAGAGGCGCAAAACGGAAA 2073  
Qy 370 AlaGlnSerProLeuAlaPheLys----- 377  
Db 2074 GCCAGTCCACCGCTGGCATTTGT-ATGAACCCCTTCCCTTGCACATTTATGTACCTTTATGA 2132  
Qy 378 ----- GluAla 379  
Db 2133 ACTCTTTATCATCATCTGAGTCTGACCATTTGATATATTTATTTCTCAACAGAAAGAGCG 2192  
Qy 380 AlaThrLysGlnLeuAlaValLeuSerSerValMetGlnCysMetGluThrHisLysLeu 399  
Db 2193 GCTACAAAGCAGCTAGCTGTGTTATCATCAGTTATGCAGTGTATGGAGACTCACAAAGTTA 2252  
Qy 400 AspProAlaLysGluLeuProGlyTyrGlnIleLysGluGlnIleValSerLeuGluLys 419  
Db 2253 GATCCTGCGAAGAACTACCCAGGATGCGAGATCAAGAGCAAAATTTGTTAGCTTGGAGAAA 2312  
Qy 420 AspThrLeuGlnLeuAspLysGluMetGluLysAlaArgSerLeuSerLeuMetGlu 439  
Db 2313 GACACTCTTCAGCTCGACAAAGAGATGGAAGAGAAAGCAAGATCTCTCAGTTTAAATGGAG 2372  
Qy 440 GluAlaAlaLeuAlaLysArgMetTyrAsnGlnIleLysArgProArgLeuSerPro 459  
Db 2373 GAAGCCGCACTTGCCAAAGAGAATGTATAACCAACAGATATAAACCCTCCAGGTTGTCAACC 2432  
Qy 460 MetGluMetProProValThrSerSerSerTyrSerProIleTyrArgAspArgSerPhe 479  
Db 2433 ATGGAAATGCCACCAAGTAACTTCTTCATCGTATCTCTATCTACCGGTAGTAGAGCTTT 2492  
Qy 480 ProSerGlnArgAspAspGlnAspGluLeuSerAlaLeuValSerSerTyrLeuGly 499  
Db 2493 CCTAGTCAAGAGAGCATGACCAAGATGAATATCAGCTCTTGTGAGTAGTACCTCGGC 2552  
Qy 500 ProSerThrSerPheProHisArgSerArgSerProGluTyrMetValProLeuPro 519  
Db 2553 CCGTCAACATCTTTTCTCTCATCGCTCAAGAGATCCCCCGGAATATATGTTCCACTTCCA 2612

Qy 520 HisGlyGlyLeuGlyArgSerValTyrAlaTyrGluHisLeuAlaProAsnSerTyrSer 539  
Db 2613 CATGGTGGTATTAGGAAGAGTATATGATATGACATCTGGCCCCCAATTCATCTCT 2672  
Qy 540 ProGlyHisGlyHisArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGlnArg 559  
Db 2673 CCAGGTCCAGGACATAGACTTCATCGACAGTACTCTCCGTCTTTGGTTTCAGGACAGAGA 2732  
Qy 560 HisProLeuGlnTyrSerProProIleHisGlyGlnGlnLeuProTyrGlyIleGln 579  
Db 2733 CATCCACTACAGTACTCTCTCCATTCATGACAAACAGTACCATATGTTATGATACAA 2792  
Qy 580 ArgValTyrArgHisSerProSerGluGluArgTyrLeuGlyLeuSerAsnGlnArgSer 599  
Db 2793 AGGGTTTACAGACATTCACCATCTCAAGAAAGATATTTGGGTTTATCCATCAAGAGTCT 2852  
Qy 600 ProArgSerAsnSerSerLeuAspProLys 609  
Db 2853 CCTCGCAGTAACATCATATTAGACCCCAAA 2892  
RESULT 3  
AAC33344  
ID AAC33344 standard; DNA; 2155 BP.  
XX AAC33344;  
AC AAC33344;  
DT 17-OCT-2000 (first entry)  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 2690.  
XX Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
OS Arabidopsis thaliana.  
XX EP1033405-A2.  
XX 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-0301439.  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125768.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130075.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.

QY 501 SerThrSerPheProHisArgSerArgSerProGluTyrMetValProLeuProHis 520  
 Db 1862 TCAACATCTTTTCCATCGCTCAAGAGATCCCGGAATATATGGTTCCACTCCACAT 1921  
 QY 521 GlyGlyLeuGlyArgSerValTyrAlaTyrGluHisLeuAlaProAsnSerTyrSerPro 540  
 Db 1922 GGTGGGTTAGGAAGAAGTGATATGTCATATGAACATCTGGCCCCAAATTCATCTCTCCA 1981  
 QY 541 GlyHisGlyHisArgLeuHisArgGluTyrSerProSerLeuValHisGlyGlnArgHis 560  
 Db 1982 GGTACCGACATAGACTTCATCGACAGTACTCTCCGTCCTTGGTTCACGACAGACAT 2041  
 QY 561 ProLeuGlnTyrSerProProIleHisGlyGlnGlnGlnLeuProTyrGlyIleGlnArg 580  
 Db 2042 CCACCTACAGTACTCTCTCCCAATTCATGGACAACAACAGTACCATATGTTATACAAAG 2101  
 QY 581 ValTyrArgHisSerProSerGluGluArgTyrLeuGlyLeuSerAsnGlnArgSerPro 600  
 Db 2102 GTTTACAGACATTCACCATCTGAAGAAGATATTGGGTTTATCCAAATCAAGGTCTCCT 2161  
 QY 601 ArgSerAsnSerSerLeuAspProlys 609  
 Db 2162 GCAGTAACCTATCATTTAGACCCCAAA 2188

## RESULT 2

ID AAA63668 standard; DNA; 3761 BP.  
 AC AAA63668;

DT 04-DEC-2000 (first entry)

DE H51 genomic sequence of first 17 kb of 84M13 encompassing FRI gene.

KW H51; one locus-FRIGIDA; FRI gene; flowering time; blotting;  
 KW flower initiation; stem elongation; flower production; ss.

OS Arabidopsis sp.

PN WO2000046359-A2.

PD 10-AUG-2000.

PF 25-JAN-2000; 2000WO-GB00197.

PR 05-FEB-1999; 99GB-0002660.

PA (PLAN-) PLANT BIOSCIENCE LTD.

PI Johanson U, West J, Dean C;

DR WPI; 2000-532899/48.

XX New nucleic acid derived from the FRI locus of a plant, e.g.  
 PT Arabidopsis, encoding a polypeptide capable of specifically altering  
 PT the flowering time of a plant.

BS Claim 4; Fig 4; 73pp; English.

CC The present sequence represents the (late flowering) H51 genomic  
 CC sequence of the first 17 kb of cosmid 84M13 encompassing FRI (one  
 CC locus-FRIGIDA) locus of Arabidopsis. The FRI gene encodes a  
 CC polypeptide capable of specifically altering the flowering time of a  
 CC plant. The FRI polynucleotide is used to transform plants, so that  
 CC the flowering time of a plant is altered. This is used, for example,  
 CC for plants in which the leaves or tubers are a commercial product,  
 CC where it is desirable to avoid 'plotting' (initiation of flowers and  
 CC stem elongation) at too early a stage. Conversely, it may be desirable  
 CC to alter flowering under certain circumstances e.g. to vary flower  
 CC production across the seasons.

XX Sequence 3761 BP; 1174 A; 717 C; 775 G; 1095 T; 0 other;

Alignment Scores:  
 Pred. No.: 3761  
 Score: 2981.00  
 Percent Similarity: 78.96%  
 Best Local Similarity: 78.96%  
 Query Match: 96.04%  
 DB: 21  
 Length: 3761  
 Matches: 608  
 Conservative: 0  
 Mismatches: 1  
 Indels: 162  
 Gaps: 2

US-09-890-475-1 (1-609) x AAA63668 (1-3761)

QY 1 MetSerAsnTyrProThrValAlaAlaGlnProThrThrThrAlaAsnProLeuLeu 20  
 Db 574 ATGTCCCAATTATCCACGACGGTGGCGGCGCAACCCACACGCGCAATCCACTGCTG 633  
 QY 21 GlnArgHisGlnSerGluGlnArgArgGluLeuProlysIleValGluThrGluSer 40  
 Db 634 CAGCGACATCATCTGAAACGACGAGAGAAATACCGAAGATTGTCGAAACAGAGTCT 693  
 QY 41 ThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLysSerIleAsp 60  
 Db 694 ACAAGTATGGACATTACGATCGGTCAATCTAAGCAGCTCAATTTTGAATCCATAGAC 753  
 QY 61 GluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAspLeuGln 80  
 Db 754 GAATTAGCTCGTTTTTCAGTTGCAGTGGAAACATTCAAACGCCAATTCGATGATCTCAG 813  
 QY 81 LysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsnGlyValVal 100  
 Db 814 AAGCACATCGAGTCAATCGAAACGCAATTGATTCCAACTCGAGAGTAACGCGGTGTC 873  
 QY 101 LeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSerProProArgAsnVal 120  
 Db 874 CTCGCCCGGGGGAACAATAATTTCCATCAGCCGATGTTATCGCTCCGCGGAACAATGTA 933  
 QY 121 SerValGluThrThrValThrValSerGlnProSerGlnGluIleValProGluThrSer 140  
 Db 934 TCTGTAGAACCCACCGTCCTGTAGCCCAACCGCTCTCAGAGAGATTGTACGGAGCGTGG 993  
 QY 141 AsnLysProGluGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyr 160  
 Db 994 AATAAACCGGGGGGACGTAATGTGTAGTTGATGTAGCAAAAGGTCTGGCTAAATAC 1053  
 QY 161 IleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGluGluIleProSerAlaLeuLys 180  
 Db 1054 ATATACCGGAATATCTCTGATCAAGCTAAGTTAATGGAAGAGATTCCTTCAGCTTGAAA 1113  
 QY 181 LeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPheTyrLeuGlnGly 200  
 Db 1114 TTGGCCAAGGAGCGACGAAAGTTTGTATTGGATTGTATTGGCAAGTTTACTTACAAGGG 1173  
 QY 201 ArgArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuIle 220  
 Db 1174 CGTAGAGCATTTACTAAAGAGTCGCTATAGAGTCCTGCGAGACAAGTTTCGGTCTTTATA 1233  
 QY 221 LeuGluSerPheLeuLeuMetProAspArgGlyLysGlyLysValLysIleGluSerTyr 240  
 Db 1234 CTGGAGTCTTTTCTTAATGCTGATCGTGGTAAAGGGAAGGTGAAGATTGAGAGTTGG 1293  
 QY 241 IleLysAspGluAlaGluThrAlaAlaValAlaIleThrArgLysArgLeuMetThrGluGly 260  
 Db 1294 ATTAAGATGAGCGGAGACGGCTGCTGTGTGTTGGAGGAAAGGTGTGATGACTGAAGA 1353  
 QY 261 GlyLeuAlaAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuValAlaCysPhe 280  
 Db 1354 GGATTAGTCGGCTGAGAAAATGGATGCAAGGGGTTTGTCTTTACTAGTTGCTGTTT 1413  
 QY 281 GlyValProSerAsnPheArgSerThrAspLeuLeuAspLeuIleArgMetSerGlySer 300  
 Db 1414 GGTCTTCTCTCAAACTTTAGAGTACAGATTTCCTGGATTGATAGAGGATGAGTGGTTCG 1473  
 QY 301 AsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetValSer----- 318

FT       /\*tag= a  
 FT       /product= "polypeptide which alters flowering time"  
 XX       WO200046358-A2.  
 XX       10-AUG-2000.  
 XX       25-JAN-2000; 2000WO-GB00197.  
 XX       05-FEB-1999; 99GB-0002660.  
 PR       (PLAN-) PLANT BIOSCIENCE LTD.  
 XX       Johanson U, West J, Dean C;  
 XX       WPI; 2000-532899/48.  
 DR       P-PSDB; AAB08030.

XX       New nucleic acid derived from the FRI locus of a plant, e.g.  
 PT       Arabidopsis, encoding a polypeptide capable of specifically altering  
 PT       the flowering time of a plant

XX       Claim 4; Fig 5; 73pp; English.

CC       The present sequence encodes a polypeptide capable of specifically  
 CC       altering the flowering time of a plant. The polypeptide is encoded  
 CC       by the FRI (one locus-FRIGIDA) locus of Arabidopsis. The FRI  
 CC       polynucleotide is used to transform plants, so that the flowering time  
 CC       of a plant is altered. This is used, for example, for plants in which  
 CC       the leaves or tubers are a commercial product, where it is desirable  
 CC       to avoid 'blotting' (initiation of flowers and stem elongation) at  
 CC       too early a stage. Conversely, it may be desirable to alter flowering  
 CC       under certain circumstances e.g. to vary flower production across the  
 CC       seasons.

XX       SQ       Sequence 2257 BP; 698 A; 443 C; 493 G; 623 T; 0 other;

#### Alignment Scores:

Pred. No.:           2,96e-244           Length:           2257  
 Score:               3104.00           Matches:           609  
 Percent Similarity: 100.00%       Conservative:       0  
 Best Local Similarity: 100.00%   Mismatch:           0  
 Query Match:       100.00%       Indels:            0  
 DB:                   21           Gaps:             0

US-09-890-475-1 (1-609) x AAA63669 (1-2257)

Qy       1   MetSerAsnTyrProThrValAlaAlaGlnProThrThrAlaAsnProLeuLeu 20  
 Db       362 ATGTCCTAATTATCCACGACGGTGGCGGCGCAACCCACACGACGCGAATCCACTGCTG 421  
 Qy       21   GlnArgHisGlnSerGluGlnArgArgGluLeuProIysIleValGluThrGluSer 40  
 Db       422 CAGCGACATCAATCTGAACGACGACGAGAGAAATTACCGAAGATTCTCGAAACAGAGTCT 481  
 Qy       41   ThrSerMetAspIleThrIleGlyGlnSerIysGlnProGlnPheLeuIysSerIleAsp 60  
 Db       482 ACAAGTATGACATTCAGATCGCTCAATCTAAGCAGCCTCAATTTTGAATCCATAGAC 541  
 Qy       61   GluLeuAlaAlaPheSerValAlaValGluThrPheIysArgGlnPheAspIleuGln 80  
 Db       542 GAATTAGTCGCTTTTTCAGTTGTCAGTGGAAACATTCAAAACGCCCAATTCGATGATCTTCAG 601  
 Qy       81   LysHisIleGluSerIleGluAsnAlaIleAspSerIysLeuGluSerAsnGlyValVal 100  
 Db       602 AAGCACATCGAGTCAATCGAAACGCAATTGATTCGAATCTCGAGAGTAAACGGCGTTGTC 661  
 Qy       101   LeuAlaAlaArgAsnAsnAsnPheHisGlnProMetIeuSerProProArgAsnVal 120  
 Db       662 CTCGCGCGCGGAACAATAATTCATTCAGCCGATGTTATCGCTCCGCGGAACAATGTA 721  
 Qy       121   SerValGluThrThrValThrValSerGlnProSerGlnGluIleValProGluThrSer 140

Db       722 TCTGTAGAAACACACCGTCACTGTGAGCCACCGTCTCAGGAGATTGTACCGGAGACGTGC 781  
 Qy       141 AsnIysProGluGlyGlyArgMetCysGluIeuMetCysSerIysGlyLeuArgIysTyr 160  
 Db       782 AATAAAACCGAGGGGAGCGTATGTGTGAGTTGATGTAGCAAAAGGTCTCGTAAATAC 841  
 Qy       161 IleTyrAlaAsnIleSerAspGlnAlaIysLeuMetGluGluIleProSerAlaLeuIys 180  
 Db       842 ATATACCGAATATCTCTGATCAAGCTAAGTTAATGAGAGAGATTCCTTCAGCTTTGAAA 901  
 Qy       181 LeuAlaIysGluProAlaIysPheValLeuAspCysIleGlyIysPheTyrIeuGlnGly 200  
 Db       902 TTGGCCAGGAGCCAGCAAGTTTCTATTGGATTGTAATGGCAAGTTTACTTCAAGGG 961  
 Qy       201 ArgArgAlaPheThrIysGluSerProMetSerSerAlaArgGlnValSerLeuLeuIle 220  
 Db       962 CGTAGAGCAATTTACTAAGAGTCGCTATGAGCTCTCGAGACAGATTTCGCTTCTATA 1021  
 Qy       221 LeuGluSerPheLeuLeuMetProAspArgGlyIysGlyIysValIysIleGluSerTyr 240  
 Db       1022 CTGGAGTCTTTTCTTAATGCTGATCGTGTAAAGGGAAGGTCAAGATTGAGAGTTGG 1081  
 Qy       241 IleIysAspGluAlaGluThrAlaAlaValAlaIleArgIysArgLeuMetThrGluGly 260  
 Db       1082 ATTAAGATGAGCGGAGACGCTCTGCTGCTTGGAGAAAAGTTGATGACTGAGGA 1141  
 Qy       261 GlyIeuAlaAlaIleGluIysMetAspAlaArgGlyLeuLeuLeuValAlaCysPhe 280  
 Db       1142 GGATTAGTCGCGCTGAGAAAATGATGCAAGGGGTTTGTCTTTTACTAGTTGCTGTTTT 1201  
 Qy       281 GlyValProSerAsnPheArgSerThrAspIeuAspIleAspMetSerGlySer 300  
 Db       1202 GGGTTTCCCTTCAAACTTAGGAGTACAGATTGCTGGATTGATAAGGATGAGTGGTTCG 1261  
 Qy       301 AsnGluIleAlaGlyAlaLeuIysArgSerGlnPheLeuValProMetValSerGlyIle 320  
 Db       1262 AATGAGATTGCGGTCGTTTGAAGCGGTCAAGATTTCTTGTCCCTATGCTCTCAGGTATA 1321  
 Qy       321 ValGluSerSerIleIysArgGlyMetHisIleGluAlaLeuGluMetValTyrPhe 340  
 Db       1322 GTTGAATCAAGTATCAAGCGTGGAAATGATGATTTGAAGCTCTTGAGATGGTTATACCTTT 1381  
 Qy       341 GlyMetGluAspIysPheSerAlaAlaLeuValLeuThrSerPheLeuIysMetSerIys 360  
 Db       1382 GGCATGGAGGATAAGTTTTCAGCTGCTCTAGTCTTAATCTTCTTAAGATGAGCAAG 1441  
 Qy       361 GluSerPheGluArgAlaIysArgIysAlaGlnSerProLeuAlaPheIysGluAlaAla 380  
 Db       1442 GAGTCATTTGAGAGGGCAAAACGGAAAGCCCAAGTCAACGCTGGCATTTTAAAGAAAGCGCT 1501  
 Qy       381 ThrIysGluLeuAlaValLeuSerSerValMetGlnCysMetGluThrHisIysLeuAsp 400  
 Db       1502 ACAACGACCTAGCTGTGTATCATCATGTTATGAGTGTATGGAGACTCACAGTTAGAT 1561  
 Qy       401 ProAlaIysGluLeuProGlyTyrGlnIleIysGluGlnIleValSerLeuGluIysAsp 420  
 Db       1562 CCTGCGAAAGAACTACCAAGTGGCAGATCAAGAGCAAAATTTGTAGCTTGGAGAAAGAC 1621  
 Qy       421 ThrIeuGluIeuAspIysGluMetGluIysAlaArgSerLeuSerLeuMetGluGlu 440  
 Db       1622 ACTCTTCAGCTCGAACAAGAGATGGAAGAGAAAGCAAGATCTCTCAGTTTAAATGGAGAA 1681  
 Qy       441 AlaAlaLeuAlaIysArgMetTyrAsnGlnGlnIleIysArgProArgLeuSerProMet 460  
 Db       1682 CGCGCACTTGGCAAGAGAAATGTATTAACCAACAGATAAAACGTCCTCAAGGTTGTACCCTATG 1741  
 Qy       461 GluMetProProValThrSerSerSerTyrSerProIleTyrArgAspArgSerPhePro 480  
 Db       1742 GAAATGCCACCAAGTAACTTCTTCATCGTATTCTCTATCTACCGGTGATAGAAGTTTCCCT 1801  
 Qy       481 SerGlnArgAspAspGlnAspGluIleSerAlaLeuValSerSerTyrIeuGlyPro 500  
 Db       1802 AGTCAAGAGACGATGACCAGATGAATATCAGCTCTTGTGAGTAGTATCTCGGCCCG 1861

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 25, 2003, 08:46:35 ; Search time 462 Seconds  
(without alignments)  
3558.350 Million cell updates/sec

Title: US-09-890-475-1

Perfect score: 3104

Sequence: 1 MSNYPPTVAAGPTTANPLL.....RYLGSLNQRSPRNSSLDPK 609

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_p2n.model -DEV=xlh  
-Q/cn2 1/USPTO spo01/US09890475/runat 14082003 090348 19534/app query.fasta\_1.775  
-DB=N Geneseq 19Jun03 -QWTF=fastcap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFWT=pcp -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09890475@CGN 1\_1\_312@runat 14082003 090348 19534 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGIOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 19Jun03.\*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*  
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3104	100.0	2257	21	CDNA sequence of t
2	2981	96.0	3761	21	H51 genomic sequen
3	349	11.2	2155	21	Arabidopsis thalia
4	340	11.0	2158	21	Arabidopsis thalia
5	296.5	9.6	1964	21	Arabidopsis thalia
6	296.5	9.6	1970	21	Arabidopsis thalia
7	288.5	9.3	1865	21	Arabidopsis thalia
8	234.5	7.6	1524	24	Arabidopsis thalia
9	152	4.9	7785	23	Arabidopsis thalia
10	151	4.9	3312	20	Drosophila melanog
11	148.5	4.8	3824	20	Arabidopsis thalia
12	148.5	4.8	4620	25	Breast cancer asso
13	148.5	4.8	4972	24	Novel human coding
14	145.5	4.7	2760	24	DNA encoding human
15	145	4.7	2187	24	Human novel polynu
16	145	4.7	2045	24	Human DNA sequence
17	144	4.6	2684	24	Human novel polynu
18	144	4.6	2756	24	Human lung cancer
19	143	4.6	3270	25	Human intracellula
20	143	4.6	18506	23	Drosophila melanog
21	142	4.6	3114	18	Human RHAMM cDNA
22	142	4.6	3114	24	Human cDNA encodi
23	141.5	4.6	4772	23	Drosophila melanog
24	141	4.5	2020	24	Human novel polynu
25	140.5	4.5	2511	22	Human polynucleoti
26	140	4.5	3718	24	Rat sequence diffe
27	137.5	4.4	6558	21	cDNA sequence of a
28	136	4.4	6455	24	Human cDNA differe
29	136	4.4	7080	23	DNA encoding novel
30	135.5	4.4	14278	23	Drosophila melanog
31	135.5	4.4	14278	23	Drosophila melanog
32	134	4.3	2526	24	Arabidopsis thalia
33	134	4.3	11528	23	Arabidopsis thalia
34	133	4.3	17612	19	Pseudomonas Xpc, O
35	133	4.3	17612	21	Pseudomonas alcali
36	133	4.3	17612	22	Pseudomonas alcali
37	133	4.3	17612	24	Pseudomonas alcali
38	132.5	4.3	3913	22	Human polynucleoti
39	132.5	4.3	5142	24	CDNA encoding huma
40	132	4.3	7903	23	Arabidopsis thalia
41	131.5	4.2	2904	23	Arabidopsis thalia
42	131	4.2	9775	20	N. crassa his-3 co
43	130.5	4.2	3397	22	Human polynucleoti
44	130.5	4.2	7393	22	Human polynucleoti
45	130	4.2	6949	23	Drosophila melanog

# ALIGNMENTS

## RESULT 1

AAA63669

ID AAA63669 standard; cDNA; 2257 BP.

XX AAA63669;

AC AAA63669;

XX 04-DEC-2000 (first entry)

DT CDNA sequence of the FRI gene of Arabidopsis.

DE H51; one locus-FRIGIDA; FRI gene; flowering time; blotting;

KW flower initiation; stem elongation; flower production; ss.

XX Arabidopsis sp.

OS Arabidopsis sp.

XX Key

XX Location/Qualifiers

XX 362...2191

DIQKHSTENAIKSLKSNGLVLAARNNHOPMLSPNNVUSVETTVTSQPSQEI  
VPETSNKPEGKMGKMLKSKIRXYIYANISDQAKLMEEIPSAKLAKPAKFVLDCl  
GRFYLQGRRAFTKESFMSSARQVSLLL"  
BASE COUNT 166 a 115 c 133 g 144 t  
ORIGIN

Alignment Scores:  
Pred. No.: 6.5e-53 Length: 558  
Score: 931.00 Matches: 185  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 99.48% Mismatches: 0  
Query Match: 29.99% Indels: 0  
DB: 8 Gaps: 0

US-09-890-475-1 (1-609) x AY092693 (1-558)

QY	34	LysIleValGluThrGluSerThrSerMetAspIleThrIleGlyGlnSerLysGlnPro	53
Db	1	AAGATTGTGGAACAGAGTCTACAAGTATGGACATTACGATCGGTCAATCTAAGCAGCCT	60
QY	54	GlnPheLeuLysSerIleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLys	73
Db	61	CAATTTTGAATCCATAGACGAATTAGCTCGGTTTTCAGTTGTCAGTGGAAACATTCAA	120
QY	74	ArgGlnPheAspAspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSerLys	93
Db	121	CGCCAAATTCGATGATATTTCAGAGCACATCGAGTCAATCGAAACGCAATTGATTCAAA	180
QY	94	LeuGluSerAsnGlyValValLeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeu	113
Db	181	CTCAGAGTAAACGGGTGTCTCGCCGCGCGGACACATTAATTTCCATCAGCCGATGTTA	240
QY	114	SerProProArgAsnAsnValSerValGluThrThrValThrValSerGlnProSerGln	133
Db	241	TCGCCCTCCCGGGAACAATGTATCTGTAGAAACCAACCGTCACTGTGAGCCACCGTCTCAG	300
QY	134	GluIleValProGluThrSerAsnLysProGluGlyArgMetCysGluLeuMetCys	153
Db	301	GAGATTGTACCGGAGACGTCCGANTAAACCGAGGGGGAGGTATGTGTGAGTTGATGTGT	360
QY	154	SerLysGlyLeuArgLysTyrIleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGlu	173
Db	361	AGCAAAGGTCTGCGTAAATACATATACGCGAATATCTCTGATCAAGCTAAAGTTAATGGAA	420
QY	174	GluIleProSerAlaLeuLysLeuAlaLysGluProAlaLysPheValLeuAspCysIle	193
Db	421	GAGATTCTTTCAGCTTTCGAATTTGCCAAGGCGGACCGAAGTTTGTATTGGATTGATT	480
QY	194	GlyLysPheTyrLeuGlnGlyArgAlaPheThrLysGluSerProMetSerSerAla	213
Db	481	GGCAAGTTTACTTACAAAGGCGTAGAGCATTTTACTTAAGAGTCGCGCTATGAGCTCTGCG	540
QY	214	ArgGlnValSerLeuLeu	219
Db	541	AGCAAGTTTCGCTCTCT	558

Search completed: August 25, 2003, 12:30:04  
Job time : 6688 secs

```

QY 174 GluileProSerAlaLeuLysLeuLysGluProAlaLysPheValLeuAspCysIle 193
Db 421 GAGATTCTCTGACGTTTGAATTTGGCCAGGAGCCAGGAAAGTTTGTATTGGATTGTATT 480
QY 194 GlyLysPheTyLeuGlnGlyArgAlaPheThrLysGluSerProMetSerSerAla 213
Db 481 GGCNAGTTTACTTACAGGGGTAGACATTTACTTAAAGAGTGCCTATGAGCTTCGGC 540
QY 214 ArgGlnValSerLeuLeu 219
Db 541 AGACAGATTTCGCTTCTT 558

RESULT 14
AY092691
LOCUS AY092691
DEFINITION Arabidopsis thaliana cultivar TSU-0 FRIGIDA protein gene, partial cds.
ACCESSION AY092691
VERSION AY092691.1 GI:20160136
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
REFERENCE 1 (bases 1 to 558)
AUTHORS Hagenblad,J. and Nordborg,M.
TITLE Sequence Variation and Haplotype Structure Surrounding the Flowering Time Locus FRI in Arabidopsis thaliana
JOURNAL Genetics 161 (1), 289-298 (2002)
MEDLINE 22013871
PUBMED 12019242
REFERENCE 2 (bases 1 to 558)
AUTHORS Hagenblad,J. and Nordborg,M.
TITLE Direct Submission
JOURNAL Submitted (27-MAR-2002) Department of Genetics, Lund University, Solvegatan 29, Lund SE-223 62, Sweden
FEATURES
    source
        1..558
            /organism="Arabidopsis thaliana"
            /mol_type="genomic DNA"
            /cultivar="TSU-0"
            /db_xref="taxon:3702"
            <1..>558
                /product="FRIGIDA protein"
                <1..>558
                    /codon_start=1
                    /product="FRIGIDA protein"
                    /protein_id="AA12144.1"
                    /db_xref="GI:20160137"
                    /translation="KIVETESTMDITIGQSKQFQFLKSIDELAAFSVAVETFKRQFD
                    DIQKHIESIENAIDSKLESNGVLAARNNNFHQMLSPNNVSVETVTSQPSQEI
                    VPETSKPGGRMCMLCSCSKGLKYIYIANSQAKLMEEIFSAKLAKAPEKFLVDCI
                    GKFLQGRATFKESPMSARQSVLL"
BASE COUNT 166 a 115 c 133 g 144 t
ORIGIN

Alignment Scores:
    Fred. No.: 6.5e-53 Length: 558
    Score: 931.00 Matches: 185
    Percent Similarity: 100.00% Conservative: 1
    Best Local Similarity: 99.46% Mismatches: 0
    Query Match: 29.99% Indels: 0
    DB: 8 Gaps: 0

US-09-890-475-1 (1-609) x AY092691 (1-558)

QY 34 LysIleValGluThrGluSerThrSerMetAspIleThrIleGlyGlnSerLysGlnPro 53
Db 1 AAGATTGTGGAACAGAGTCTACAGTAGTGACATTACGATCGGTCAATCTAAGCACCT 60
QY 54 GlnPheLysSerIleAspGlnLeuAlaPheSerValAlaValGluThrPheLys 73

```

```

Db 61 CAATTTTGAATCCATAGACGAATTAGCTGCGGTTTTCAGTTCAGTGGAAACATTCAAA 120
QY 74 ArgGlnPheAspLeuGlnLysHileGluSerIleGluAsnAlaIleAspSerLys 93
Db 121 CGCCAAATTCGATGATATTTCAGAACACATCGAGTCAATCGAAACGCAATTGATTCAAA 180
QY 94 LeuGluSerAsnGlyValValLeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeu 113
Db 181 CTCGAGAGTAACGGCGTTGTCTCCGCGCGGCAACAATAATTTCATCAGCCGATGTTA 240
QY 114 SerProProArgAsnValSerValGluThrThrValThrValSerGlnProSerGln 133
Db 241 TCGCCTCCGCGGACAAATGATCTGTAGAACACACCGCTCACTGTGACCAACCGTCTCAG 300
QY 134 GluIleValProGluThrSerAsnLysProGluGlyArgMetCysGluLeuMetCys 153
Db 301 GAGATGTGTACCGGAGCGTCGAATAAACCGGAGGGGCGATGTGTGTGATTTGATGTGT 360
QY 154 SerLysGlyLeuArgLysTyIleTyAlaAsnIleSerAspGlnAlaLysLeuMetGlu 173
Db 361 AGCAAGGCTCTGGTAATATACATATACGGAATATCTCTGATCAAGCTAAGTTAATGGAA 420
QY 174 GluileProSerAlaLeuLysLeuAlaLysGluProAlaLysPheValLeuAspCysIle 193
Db 421 GAGATTCTCTGACGTTTGAATTTGGCCAGGAGCCAGGAAAGTTTGTATTGGATTGTATT 480
QY 194 GlyLysPheTyLeuGlnGlyArgAlaPheThrLysGluSerProMetSerSerAla 213
Db 481 GGCNAGTTTACTTACAGGGGTAGACATTTACTTAAAGAGTGCCTATGAGCTTCGGC 540
QY 214 ArgGlnValSerLeuLeu 219
Db 541 AGACAGATTTCGCTTCTT 558

RESULT 15
AY092693
LOCUS AY092693
DEFINITION Arabidopsis thaliana cultivar Ler FRIGIDA protein gene, partial cds.
ACCESSION AY092693
VERSION AY092693.1 GI:20160140
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
REFERENCE 1 (bases 1 to 558)
AUTHORS Hagenblad,J. and Nordborg,M.
TITLE Sequence Variation and Haplotype Structure Surrounding the Flowering Time Locus FRI in Arabidopsis thaliana
JOURNAL Genetics 161 (1), 289-298 (2002)
MEDLINE 22013871
PUBMED 12019242
REFERENCE 2 (bases 1 to 558)
AUTHORS Hagenblad,J. and Nordborg,M.
TITLE Direct Submission
JOURNAL Submitted (27-MAR-2002) Department of Genetics, Lund University, Solvegatan 29, Lund SE-223 62, Sweden
FEATURES
    source
        1..558
            /organism="Arabidopsis thaliana"
            /mol_type="genomic DNA"
            /cultivar="Ler"
            /db_xref="taxon:3702"
            <1..>558
                /product="FRIGIDA protein"
                <1..>558
                    /codon_start=1
                    /product="FRIGIDA protein"
                    /protein_id="AA12146.1"
                    /db_xref="GI:20160141"
                    /translation="KIVETESTMDITIGQSKQFQFLKSIDELAAFSVAVETFKRQFD

```

CDS  
/product="FRIGIDA protein"  
<1..>558  
/codon\_start=1  
/product="FRIGIDA protein"  
/protein\_id="AA012134.1"  
/db\_xref="GI:20160117"

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
22013871  
12019242  
2 (bases 1 to 558)  
Hagenblad, J. and Nordborg, M.  
Sequence Variation and Haplotype Structure Surrounding the  
Flowering Time Locus FRI in Arabidopsis thaliana  
Genetics 161 (1), 289-298 (2002)

BASE COUNT 166 a 115 c 133 g 144 t  
ORIGIN

Alignment Scores:  
Pred. No.: 6.5e-53 Length: 558  
Score: 931.00 Matches: 185  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 99.46% Mismatches: 0  
Query Match: 29.99% Indels: 0  
DB: 8 Gaps: 0

US-09-890-475-1 (1-609) x AY092681 (1-558)

QY 34 LysileValGluThrGluSerThrSerMetAspIleThrIleGlnSerLysGlnPro 53  
Db 1 AAGATTGTCGAAACAGAGCTCAGAGTATGACATTCAGTACGTCGTCATTAAGCAGCT 60  
QY 54 GlnPheLeuLysSerIleAspGluLeuAlaPheSerValAlaValGluThrPheLys 73  
Db 61 CAATTTTGAATCCATAGACGAATTAGTCGCTTTTCACTGAGTGGAAACATTCAAA 120  
QY 74 ArgGlnPheAspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSerLys 93  
Db 121 CGCCAAATTCGATGATATTCAGAACACATCAGTCAATCGAAACGCAATTCATCCAA 180  
QY 94 LeuGluSerAsnGlyValValLeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeu 113  
Db 181 CTCGAGAGTAAACGGCGTTGCTCTCCGCGCGGAGCGATGTCGTGAGTGAATGTTA 240  
QY 114 SerProProAsnAsnValSerValGluThrValThrValSerGlnProSerGln 133  
Db 241 TCGCCTCCGCGGAGCAATGATCTGTGAACACCACTGAGTGAACCAACCTCTCAG 300  
QY 134 GluileValProGluThrSerAsnLysProGluGlyArgMetCysGluLeuMetCys 153  
Db 301 GAGATTGTCGAGAGAGCTGCAATAAACCAGCGAGCGGAGCGATGTCGTGAGTGAATG 360  
QY 184 SerLysGlyLeuArgLysTyrlleTyAlaAsnIleSerAspGlnAlaLysLeuMetGlu 173  
Db 361 AGCAAGGTCGCGTAAATACATATACGCAATATCTCTGATCAAGCTAAGTTAATGGAA 420  
QY 174 GluileProSerAlaLeuLysLeuAlaLysGluProAlaLysPheValLeuAspCysIle 193  
Db 421 GAGATTCTTCAGCTTTGAATTCGCCAGAGGAGCGAGGATTTGATTTGATTTGAT 480  
QY 194 GlyLysPheTyrlleGlnGlyArgArgAlaPheThrLysGluSerProMetSerSerAla 213  
Db 481 GCGCAGTTTACTTACAGGCGTAGACATTTACTAAGAGTCGCTAGCTGCG 540  
QY 214 ArgGlnValSerLeuLeu 219  
Db 541 AGACAAAGTTTCGCTTCTT 558

RESULT 13  
LOCUS AY092683  
DEFINITION Arabidopsis thaliana cultivar GOT-32 FRIGIDA protein gene, partial cds  
ACCESSION AY092683  
VERSION AY092683.1 GI:20160120  
KEYWORDS Arabidopsis thaliana (thale cress)  
SOURCE Arabidopsis thaliana  
ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
22013871  
12019242  
2 (bases 1 to 558)  
Hagenblad, J. and Nordborg, M.  
Sequence Variation and Haplotype Structure Surrounding the  
Flowering Time Locus FRI in Arabidopsis thaliana  
Genetics 161 (1), 289-298 (2002)

BASE COUNT 166 a 115 c 133 g 144 t  
ORIGIN

Alignment Scores:  
Pred. No.: 6.5e-53 Length: 558  
Score: 931.00 Matches: 185  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 99.46% Mismatches: 0  
Query Match: 29.99% Indels: 0  
DB: 8 Gaps: 0

US-09-890-475-1 (1-609) x AY092683 (1-558)

QY 34 LysileValGluThrGluSerThrSerMetAspIleThrIleGlnSerLysGlnPro 53  
Db 1 AAGATTGTCGAAACAGAGCTCAGAGTATGACATTCAGTACGTCGTCATTAAGCAGCT 60  
QY 54 GlnPheLeuLysSerIleAspGluLeuAlaAlaPheSerValAlaValGluThrPheLys 73  
Db 61 CAATTTTGAATCCATAGACGAATTAGTCGCTTTTCACTGAGTGGAAACATTCAAA 120  
QY 74 ArgGlnPheAspLeuGlnLysHisIleGluSerIleGluAsnAlaIleAspSerLys 93  
Db 121 CGCCAAATTCGATGATATTCAGAACACATCAGTCAATCGAAACGCAATTCATCCAA 180  
QY 94 LeuGluSerAsnGlyValValLeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeu 113  
Db 181 CTCGAGAGTAAACGGCGTTGCTCTCCGCGCGGAGCGATGTCGTGAGTGAATGTTA 240  
QY 114 SerProProAsnAsnValSerValGluThrValThrValSerGlnProSerGln 133  
Db 241 TCGCCTCCGCGGAGCAATGATCTGTGAACACCACTGAGTGAACCAACCTCTCAG 300  
QY 134 GluileValProGluThrSerAsnLysProGluGlyArgMetCysGluLeuMetCys 153  
Db 301 GAGATTGTCGAGAGAGCTGCAATAAACCAGCGAGCGGAGCGATGTCGTGAGTGAATG 360  
QY 184 SerLysGlyLeuArgLysTyrlleTyAlaAsnIleSerAspGlnAlaLysLeuMetGlu 173  
Db 361 AGCAAGGTCGCGTAAATACATATACGCAATATCTCTGATCAAGCTAAGTTAATGGAA 420  
QY 174 GluileProSerAlaLeuLysLeuAlaLysGluProAlaLysPheValLeuAspCysIle 193  
Db 421 GAGATTCTTCAGCTTTGAATTCGCCAGAGGAGCGAGGATTTGATTTGATTTGAT 480  
QY 194 GlyLysPheTyrlleGlnGlyArgArgAlaPheThrLysGluSerProMetSerSerAla 213  
Db 481 GCGCAGTTTACTTACAGGCGTAGACATTTACTAAGAGTCGCTAGCTGCG 540  
QY 214 ArgGlnValSerLeuLeu 219  
Db 541 AGACAAAGTTTCGCTTCTT 558

RESULT 13  
LOCUS AY092683  
DEFINITION Arabidopsis thaliana cultivar GOT-32 FRIGIDA protein gene, partial cds  
ACCESSION AY092683  
VERSION AY092683.1 GI:20160120  
KEYWORDS Arabidopsis thaliana (thale cress)  
SOURCE Arabidopsis thaliana  
ORGANISM Arabidopsis thaliana



```

intron 11019..11078
/gene="AT4g00450"
/number=3
exon 11079..11423
/gene="AT4g00450"
/number=4
intron 11424..11474
/gene="AT4g00450"
/number=5
exon 11475..13674
/gene="AT4g00450"
/number=6
intron 13675..13734
/gene="AT4g00450"
/number=7
exon 13735..13823
/gene="AT4g00450"
/number=8
intron 13824..13920
/gene="AT4g00450"
/number=9
exon 13921..16530
/gene="AT4g00450"
/number=10
intron 16531..16676
/gene="AT4g00450"
/number=11
exon 16677..16862
/gene="AT4g00450"
/number=12
intron 16863..16975
/gene="AT4g00450"
/number=13
exon 16976..17042
/gene="AT4g00450"
/number=14
intron 17043..17091
/gene="AT4g00450"
/number=15
exon 17092..17547
/gene="AT4g00450"
/number=16
gene 18027..19004
/gene="AT4g00460"
complement(join(18027..18146,18234..18578,18669..19004))
/gene="AT4g00460"
CDS complement(join(18027..18146,18234..18578,18669..19004))
/note="similarity to
contains EST gb:AI993831.1"
/codon_start=1
/product="putative protein"
/protein_id="CAB80855.1"
/db_xref="GI:7267385"
/translation="MVTSVDFDPCLLQELADSDSFVFAEKGSLSMKARSSTGSFR
KIVQREBKWLPVLPVPSGSLDKARKQKLNKRNSTQIHKAAVAINSSILSMEI
PDSYMTLPKCGSSVDSIYRYMSGSRFFPEQLDCLNISEHEAVQLADRVASM
YWRKSKLSNKSNNWMDLMTSTTRTDKNVMAERATLLFCLKQRYPELSQSL
DICKIOYKDVGVAVLESYSRVLEGLAFNIVAWIDVLYVDKTMRGSE"
complement(18027..18146)
/gene="AT4g00460"
exon 18027..18146
/number=1
intron complement(18147..18233)
/gene="AT4g00460"
exon complement(18234..18578)
/gene="AT4g00460"
/number=2
intron complement(18579..18668)
/gene="AT4g00460"
/number=3
exon complement(18669..19004)
/gene="AT4g00460"

```

```

/number=3
19070..19781
/gene="AT4g00470"
complement(join(19070..19221,19270..19386,19472..19781))
/gene="AT4g00470"
CDS complement(join(19070..19221,19270..19386,19472..19781))
/gene="AT4g00470"
/note="similarity to
contains EST gb:AA721856"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB80856.1"
/db_xref="GI:7267386"
/translation="MENISNFDENDHQSRSIDQNDQSAVETVYSTMIDSEVYPR
TCSETTGSPDQIDETNSFCSEASPCNWPVLTESKSKCLSGLEMQNECLVQGEISE
PELEWKEKFAKLLGEDMSGSGKVCATVITISNAITLWYDNCNKNDISFDNVCYSYVW
TSEVAVRNRESISVERNELSFICMLHCR1"
complement(19070..19221)
/gene="AT4g00470"
/number=1
exon

Alignment Scores:
Pred. No.: 1,08e-181 Length: 197975
Score: 2836.00 Matches: 603
Percent Similarity: 78.95% Conservative: 1
Best Local Similarity: 78.82% Mismatches: 5
Query Match: 94.59% Indels: 158
DB: 8 Gaps: 2
US-09-890-475-1 (1-609) x ATCHRIV2 (1-197975)
Qy 1 MetSerAsnTy:ProProThrValAlaAlaGlnProThrThrThrAlaAsnProLeuLeu 20
Db 75713 ATGTCCAATTATCCACCGCGGTGGCGGCAACCAACAGACGCGGAATCCACTGCTG 75772
Qy 21 GlnArgHisGluSerGluGlnArgArgGluLeuProLysIleValGluThrGluSer 40
Db 75773 CAGGACATCAATCTGAACAGCGACGAGAGAGATTACCGAGATTGTGGAACAGAGTCT 75832
Qy 41 ThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLysSerIleAsp 60
Db 75833 ACAAGTATGGACATTACGATCGGTCAATTAAGACGCGCTCAATTTTGAATCCATAGAC 75892
Qy 61 GluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAspAspLeuGln 80
Db 75893 GAATTAGTCGGTTTCAGTTCAGTGGAAACATTCAAACGCCAATTCGATGATCTTCAG 75952
Qy 81 LysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsnGlyValVal 100
Db 75953 AAGCACATCGAGTCAATCGAAAACGCAATTGATTCCAAACTCGAGAGTAACGGCGTTGTC 76012
Qy 101 LeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSerProProArgAsnVal 120
Db 76013 CTGCGCGCGGGAACAAATTAATTCATGACCGCATGTTATGCGCTCCCGGGAACATGTA 76072
Qy 121 SerValGluThrThrValThrValSerGlnProSerGlnGluIleValProGluThrSer 140
Db 76073 TCTGTAGAACCAACCGCTCACTGTGAGCCCAACCGCTCTCAGGAGATTGTACCAGAGACGTCG 76132
Qy 141 AsnLysProGluGlyGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyr 160
Db 76133 AATAAACCGGAGGGGGAACGTTATATGTGAGTTGATGTGTAGCAAGGTTCTCGTAATAATAC 76192
Qy 161 IleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGluGluIleProSerAlaLeuLys 180
Db 76193 ATATACGGAATATCTCTGATCAAGCTAAGTTAATGGAAGAGATTCTTCAGCTTTGAA 76252
Qy 181 LeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPheTyrIleuGlnGly 200
Db 76253 TTGCCAAGGAGCGACGGAAGTTTGTATTGGATTGATTGGCAAGTTTACTTCAAGGG 76312
Qy 201 ArgArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuIle 220

```

```
QY 445 LysArgMetTyrAsnGlnGlnIleLysArgProArgLeuSerProMetGluMetProPro 464
D 58637 AAGAGATGATTAACCAACAGATTAACCAAGTCAAGTATATCAATCCATGAATGCCACCA 58696
QY 465 ValThrSerSerTyrSerProIleTyrArgAspArgSerPheProSerGlnArgAsp 484
D 58697 GTAACATCTCTCATCGTATCTCTCATCTACCGTATAGAGAGCTTTCTCTAGTCAAAGAGAC 58756
QY 485 AspAspGlnAspGluLeuSerAlaLeuValSerSerTyrLeuGlyProSerThrSerPhe 504
D 58757 GATGACCAAGATGAATATACAGCTCTGTGAGTAGTATACCTCGGCCCGCTCAACATCTTT 58816
QY 505 ProHisArgSerArgSerProGlnTyrMetValProLeuProHisGlyGlyLeuGly 524
D 58817 CCTCATCGCTCAAGAAGATCCCGGAATATATGTTCCATCTCCACATGGTGGGTAGGA 58876
QY 525 ArgSerValTyrAlaTyrGluHisLeuAlaProAsnSerTyrSerProGlyHisGlyHis 544
D 58877 AGAAGTGTATATGCATATGAACATCTGCGCCCAATTCATATCTCCAGGTACCGGACAT 58936
QY 545 ArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGlnArgHisProLeuGlnTyr 564
D 58937 AGACTTCATCGACAGTACTCTCGCTTGTGTTACGGACAGAGACATCCACTACAGTAC 58996
QY 565 SerProProIleHisGlyGlnGlnGlnLeuProTyrGlyIleGlnArgValTyrArgHis 584
D 58997 TCTCCCTCCAAATCATGACCAACACAGTATACCATATGTTATATCAAAAGGTTTACAGACAT 59056
QY 585 SerProSerGluGluArgTyrLeuGlyLeuSerAsnGlnArgSerProArgSerAsnSer 604
D 59057 TCACCATCTGAGAAGAGATATTGGGTTATCCATCAAAAGTCTCTCCGACGTAACATCA 59116
QY 605 SerLeuAspProLys 609
D 59117 TCATTAGACCCCAAA 59131

RESULT 11
LOCUS ATCHRIV2 197975 bp DNA linear PLN 16-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 2.
ACCESSION AL161472
VERSION AL161472.2 GI:7267383
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 197975)
AUTHORS Lamar, B., Stoneking, T., Stumpf, J., Mewes, H.W., Lemcke, K. and
Mayer, K.P.X.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 197975)
AUTHORS EU Arabidopsis sequencing project.
TITLE Direct Submission
JOURNAL Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bbsrc.ac.uk
COMMENT Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
this fragment has an overlap with ATCHRIV1 at the 5' end and an
overlap with ATCHRIV3 at the 3' end.
FEATURES
source
1..197975
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/variety="Columbia"
/db_xref="taxon:3702"
```

```
/chromosome="4"
10352..17547
/gene="AT4g00450"
join(10352..10384,10432..10545,10605..10709,10849..11018,
11079..11423,11475..13674,13735..13823,13921..16530,
16677..16862,16976..17042,17092..17547)
/gene="AT4g00450"
/note="Contains aminoacyl-transfer RNA synthetases
class-II signatures AA1343-1352; Prokaryotic membrane
lipoprotein lipid attachment site AA349-359
contains EST gb:AI999580.1"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB80854.1"
/db_xref="GI:7267384"
/translacion="MLSPLFVISTGAVRCKLRAINESRALKRVKIKELSDSNPQPIF
CAEPRAGVYGLPSGLLCKAPFFEQRSCEETKRWIESLQOHRKRLADNPFG
YRRKTLFVLRNNVPLLRATWFFKVTYLNQVHCWAINCDLVLLLPQHLTRHKL
GVSNQKMLNLCNTSMNFCGIAHLLSKLEIGHRCFIQPCKEIIVQOQAFAR
KELTLNGGICVFSYGTMLKGFLLISLTIGFSSYSFVNFHALQEKIEIFELQ
LLPIVYGLSIVLSQTVQSLVAIVRFCEPAPGSDLVNDSRRAYTISALIEVR
YLVAAPDTVASDFFPLPSPAACGNDVYTSKAVENLEKLSRAEISAEIQAQGRG
VLSREFTSFDTYTTICRSADDLAKIASAGYQHNVAQVQALDKALDGDIDRAVS
VLFEDLCNGADENITDVSPLASSLRWIGAISTSVCSVFLEIHWATCDFRDFRAG
VKOTKCGRKDCSQVILVLIQLKQKILGGEFTARKGNKNMFLGVKSGSDMDAPE
SPGLHDIIIVCIDOEHVHGKAKRLQLVLELIRSGIFNPIAYVQLIIVSGMIDVIQ
PAVPERMRHRIILKQLPGCFVHETLEEAQLFGGDKLSEAVRTYSNERLLRELV
EKGYNNLNLSDQSKISTSLSSVFPACNAKSKGPRKTKSSVDIRELKERI
SALLQPMSCGVETPVDRFONSKVSGSVYSKMDPEATPCEDCRBRKPMND
EKSSCYQNSPIASDEENMWIKGSKTVSSSLKVDPIETIKOVRGRGOKMARKTQS
LAQQAARLESGQASTSHVCDNKVSCPHGPGVEGENQKVDVFTSTFVSVGN
SLKQLQVKKRSIAVWLITAVRQLVSEPKSVRVGQFNKGAPVEBKNTIRWKLGADE
LYSLFLDILDLVSAAGRNLTVPNNMNCCEIGEAIVLSLRRYENILLASDLV
PEAMTALNRAASLMSNGKISGSAALVTRYILKRYSLSPVVEHNNFKATSEKL
LSELDRSGNGEYGNLPGVAGVNDPDIYLRKISIGGARSPRVNDRDLQRHVE
EATHYLKLIQGTWKASLAEKNDGQVAAQIVVGLMDCIROTGAAGBDFSLVSS
VSAIINSGVLSVARTDFSLGNIYONHPGSDSSNIARVLIHITCLLKEALGE
QSVVFELALATESLTALTVFAPVKSRGQHLSPESYDSNANNSTIDMSNGTQMA
LSRAKTAIVASVAVIGISTHGVITLRIKIVGLLRKDYLDFQVFAVKSSNGSRA
MGAKRSFPIVYVHVRFLVGNCTVSEGLVLELVGESSVAISQRMMLPLKLVFP
PAYISITVLRWPFVNSNSNSVHEDTHRYLSLTMAFHVIKHPFDVCFRDYQG
LYEIVADSTDABFASVESHGDLMLKSVAFAPLARLELNSLIDCKVPSSGYHML
VSEAKNHQNGTKLVKLVSLDCLQAPKPMQWVLELLLEQALAEKLNHMLPL
TDAIRSCPTSEKPDASENEKNFTQILLTLVLRPOAVPLFSEVHLFGESVDSMLK
QAEWFLAGQDVLFGKTIROKLIIVGESKGLTPKQPKWPGWGNSSSSSHITANKAG
KCKRFETISEEGEVEIEGSRKVLUPRVLDENSPSVGYGITERAFVOLLPCIDQ
SSDSRSTFVNLVRFNSNIEQQLSVNRSSTSNQMGTAGSSGSEISNKGSTRKL
RGSPSLARRSSANTDTSPPPAALRASMLRQLFLRLPLVIPCSPFKNTKHAL
ACTVRLIGSRVYEDYAVCSPRSELKSTESTIDPSSMADLSSEVLPFRRLTGLS
TNFKSLKVLGCKNELSRMOLPDTIRWIOAAMPILLPLSLRSLSCOPHSVPPTAL
TLVOPSGSTAAAGTNQNSPAISKSGTAAAGKIKPTMLAPHQOQEQADNTDVPDPTL
LEDGTSGLSSNASNSDANLRATCWLKGVAVRVRTDITYVGSVDDDS"
10352..10384
/gene="AT4g00450"
/number=1
10385..10431
/gene="AT4g00450"
/number=1
10432..10545
/gene="AT4g00450"
/number=2
10546..10604
/gene="AT4g00450"
/number=2
10605..10709
/gene="AT4g00450"
/number=3
10710..10848
/gene="AT4g00450"
/number=3
10849..11018
/gene="AT4g00450"
/number=4
```

```

/ gene="F6N23.19"
Complement(join(22410..22521,23220..23538,23620..23939,
24009..24137,24240..24352,24439..24548,24639..24753))
/ gene="F6N23.19"
/codon_start=1
/evidence=not experimental
/protein_id="AA013637.1"
/db_xref="GI:3047126"
/translation="MSPSPYINVLKPPFPHLMDQDILTNVVKYGRNWKRIAECLPG
SENRNDVOCQRHLKVLDPDLQKGMKEDELSELVDKYMENDPPWKSISKEL
PGRIGKCRNRHNLNPTIIKSPWREBELILVQQRGNKNAEIAKLPLPRTENN
IKHWKCVKRLLEQFNLGVVYKSPSSGFYNFNQNTNWESCITTSQIKBA
KSPQRLDLITGLNWRSSSTSLRGEESVSDVSCARLNACTLPQNSNNDVA
CKVEKMERLMAARTFTPSIIISKTSPASGLKRLQKDYDTPFPDARSHMSSE
DVSASPSKSYFVKRNTCSGKPLERRLDPDFULWDBHGRNGIVNFSVRIILPKS
DLKSLVRFPLR"
Alignment Scores:
Pred. NO.: 3.93e-182 Length: 91849
Score: 2936.00 Matches: 603
Percent Similarity: 78.95% Conservative: 1
Best Local Similarity: 78.82% Mismatches: 5
Query Match: 94.59% Indels: 158
DB: 8 Gaps: 2
US-09-890-475-1 (1-609) x F6N23 (1-91849)
QY 1 MetSerAsnTyrProThrValAlaAlaGlnProThrThrAlaAsnProLeuLeu 20
DB 56839 ATGTCCAATTATCCACGCGGTGGCGGCGCAACCCACACGACGCGGAATCCACTGCTG 56898
QY 21 GlnArgHisGlnSerGluGlnArgArgGluLeuProLysIleValGluThrGluSer 40
DB 56899 CAGCGACATCAATCTGAACGCGAGCGAAGAGATTACCGAAGATTGTCGAACAGAGTCT 56958
QY 41 ThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLysSerIleAsp 60
DB 56959 ACAAGTATGGACATTAGATCGGTCAATCTAAGCAGCCTCAATTTTGAATCCATAGAC 57018
QY 61 GluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAspAspLeuGln 80
DB 57019 GAATTAGTGGCTTTTCAGTTGAGTGGGAACATTCAAACGCGCAATTCGATGATCTTCAG 57078
QY 81 LysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsnGlyValVal 100
DB 57079 AAGCACATCGAGTCAATCGAAACGCAATTGATTCCAAACCTCAGAGATACGCGCTGTGC 57138
QY 101 LeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSerProProArgAsnAsnVal 120
DB 57139 CTCGCGCGCGGAACAATAATTTCCATCAGCCGATGTTATCGCCTCCGCGGAACAATGTA 57198
QY 121 SerValGluThrThrValThrValSerGlnProSerGlnGluIleValProGluThrSer 140
DB 57199 TCTGTAGAAACACCGTCACCTGTGAGCGCAACCGTCTCAGAGAGATTGTACCGAGAGCTGC 57258
QY 141 AsnLysProGluGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyr 160
DB 57259 AATAAACCGGAGGGGAACGATATATGTAGTTGATGTAGTAGCAAGAGTCTGCGTAAATAC 57318
QY 161 IleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGluGluIleProSerAlaLeuLys 180
DB 57319 ATATACGCGAATACTCTGTATCAAGCTAAGTTAATGGAAGAGATTCTTCAGCTTTGAAA 57378
QY 181 LeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPheTyrLeuGlnGly 200
DB 57379 TTGGCCAAAGGAGCGAGCAAGTTTGTATTGATTTGATTTGGCAAGTTTACTTACAAGG 57438
QY 201 ArgArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuIle 220
DB 57439 CGTAGAGCAATTATCTAAAGAGTCGCCATGAGTCTGCGGAGCAAGTTTCGCTCTCTATA 57498
QY 221 LeuGluSerPheLeuLeuMetProAspArgGlyLysGlyLysValLysIleGluSerTyr 240

```

```

DB 57499 CTGAGCTCTTTCTTCTTAATGCCTCATCGTGTAAAGGAAAGGTCAAGATTGAGATTGG 57558
QY 241 IleLysAspGluAlaGluThrAlaAlaValAlaIleTyrArgLysArgLeuMetThrGluGly 260
DB 57559 ATTAAGATGAGCGGAGACGGCTCTGTTGCTTGGAGAAAAGGTGATGACTGAAGGA 57618
QY 261 GlyLeuAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuLeuValAlaCysPhe 280
DB 57619 GGATTAGCTCGGCTGAGAAAAATGATGCAAGGGGTTTGTCTTTACTAGTTGCTGTITT 57678
QY 281 GlyValProSerAsnPheArgSerThrAspLeuLeuAspLeuIleArgMetSerGlySer 300
DB 57679 GGTGTTCTCTCAACCTTAGGAGTACAGATTCTGGATTTGATGAAGATGAGTGTTCG 57738
QY 301 AsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMet 316
DB 57739 AATGAGATTGCCGGTCTTTGAAGCGGTCAAGTTTCAGGTACCATATTCCTGTTCTCAC 57797
QY 316 ----- 316
DB 57798 TCGTGAAATTTCAATGCAAAAGGTGTTCTTTTGTGTGACATCATCGACCAACATCAAGTT 57857
QY 316 ----- 316
DB 57858 CCATCTTGTGTTTTCGATAGCTTGTATGATATAAACTAGGAGACACATCAATATTATTAG 57917
QY 316 ----- 316
DB 57918 AGTGCAATGACTGATGACCCAAATCTTAGCTAGAGAATAATCTGGAAAGAACTTGGAAC 57977
QY 316 ----- 316
DB 57978 TCTCAACCATAGCTTTTGGTAGCAAAATGTTGCTTGTTCAGAACCAAAATGATAGCTATTG 58037
QY 316 ----- 316
DB 58038 CCTTGAATAGTGTCTTCTTGTGTTTCCAATATTGGAAGTTAAAAATCATATGACTTAGCT 58097
QY 316 ----- 316
DB 58098 GTTGGATCTAATTATAGCTTAAGCAATGCCAATCTAAGAAAGTGGTACTTACACAATATT 58157
QY 317 -----ValSerGlyIleValGluSerSerIleLysArgGlyMetHisIleGluAlaLeu 334
DB 58158 CTATTGCTCATAGTATAGTTGAATCAAGTATCAAGCGTGAATGCATATTGGAAGCTCTT 58217
QY 335 GluMetValTyrThrPheGlyMetGluAspLysPheSerAlaAlaLeuValLeuThrSer 354
DB 58218 GAAATGTTTATACCTTTTGGCATGGAGGATAGTTTTCAGCTGCTCTAGTTCTTAACTTCA 58277
QY 355 PheLeuLysMetSerLysGluSerPheGluArgAlaLysArgLysAlaGlnSerProLeu 374
DB 58278 TTCTTAAAGATGAGCAAGAGTCAATTGAGAGGGCAAAACGMAAGCCAGTACCCTG 58337
QY 375 AlaPheLys----- 377
DB 58338 GCATTTGT-ATGAACCCCTTCCTTGCACATATGATGACCTTTATGAACCTCTTTATCATCAT 58396
QY 378 -----GluAlaAlaThrLysGlnLeu 384
DB 58397 CTGAGCTGACCATTCATATATTTTCTCAACAGAAAGAGCGCTACAAAGCAGCTA 58456
QY 385 AlaValLeuSerSerValMetGlnCysMetGluThrHisLysLeuAspProAlaLysGlu 404
DB 58457 GCTGTGTATCATCATGATTATGAGTGTATGGAGATCTCAAGTGTAGTCTTCGCGAAGAA 58516
QY 405 LeuProGlyTyrGlnIleLysGluGlnIleValSerLeuGluLysAspThrLeuGlnLeu 424
DB 58517 CTACGAGATGGCAGATCAAGAGCAAAATTTGTAGCTTGGAGAAAGACACTCTTCAGTCT 58576
QY 425 AspLysGluMetGluGluLysAlaArgSerLeuSerLeuMetGluGluAlaAlaLeuAla 444
DB 58577 GACAAAGAGATGGAAGAAAGCAAGATCTCTCAGTTTAAATGGAGAAAGCCGCTTCC 58636

```

by M. Marra, WashU, to pick the best candidates for sequencing.

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

#### NEFBORING COSMID INFORMATION:

The 5' clone is F5110, 200 bp overlap; 3' clone is F15P23, 900 bp overlap. Actual start of this clone is at base position 104960 of CELF5110; actual end is at 91149 of CELF6N23

The clone sequenced to the left is F5110. The actual start of this clone is at base position 104960 of F5110; actual end is at base position 91149 of F6N23.

#### NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder (P. Green and L. Hillier, ms in preparation).

#### FEATURES

```

source
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="IV"
/map="unknown"
/clone="F6N23"
complement(187..907)
/gene="F6N23.23"
complement(join(187..347,396..512,598..907))
/gene="F6N23.23"
/codon_start=1
/evidence=not experimental
/protein_id="AAC13624.2"
/db_xref="GI:6382044"
/translation="MENLSNPDENDDHQSPRIDNQSAVETPVYTMGIDFVYPR
TCSETSGFSQIDETNSFCSEAFPCNPVLTESKSKCLSGLEMQNECLVQEISE
PELETWKEKFAKLLGEDMSGSGKGVCTAVTISNAITNLVDCNKDTSFONVCSYVW
TESEVRVAVNREESIVERNELSFICMLHCRHP"
complement(4919..7497)
/gene="F6N23.22"
complement(join(4919..4993,5078..5488,5567..6055,
6157..6213,6498..6594,6935..7222,7349..7497))
/gene="F6N23.22"
/codon_start=1
/evidence=not experimental
/protein_id="AAC13623.1"
/db_xref="GI:3047112"
/translation="MSLWADGVEAAGRSKQNSLLRKQLALAVRQVMSYAFWSS
SUTQGVLEWGGYGNMKKKYSYSHYKYGKSKELRKLKYLMLSGDSGTST
TDNLNDNDNDNDKSLSPDLSDEEYLVMSYVFSQCLPGRASATGETTTL
CNAQVNDKLFERSLIARTVCFYPIGVIELGTETISDNLRLNIKSLMEISAH
QNDDEKWEIKISEKQLPLGISDEDLHKRTISTVLNYSADRSGKNDKNRHRQP
NIVTSEPGSSFLRMKQSQVGFQKKSONVLRKILHDVPLMKRMPSONSGLN
QDPSRKEKEKESVLTMTVNEVDKESILNNTIKYLQEARVEELESOMGSVN
FYERQKTNLNDVLIETFSGNVDDSKIDNSGETQVTPROKTHLRKVKETE
VIVYRCSYRDVIVADIMETLSNLHWDAPSFVSRHTLNKFLTNLKAKEFGAAVSQVM
IKRELRRVIGDLF"
10235..12675
/gene="F6N23.1"
join(10235..10495,10693..10896,10998..11192,11341..11505,

```

```

11633..11893,11998..12216,12315..12534,12617..12675)
/gene="F6N23.1"
/Note="similar to the family of glycosyl hydrolases"
/evidence=not experimental
/protein_id="AAC13634.1"
/db_xref="GI:3047123"
/translation="MAIRLNSHVIYPSVKLGAPTRVSARSSLPFSVGDWRGVSTFSGA
RPLVLAKVLRABSTEEDRVPIDDDDSDQLVSVFHLDDSLGLVDMNSEVVEPEE
LLDLRLTKSVNVGVNDCWMIQVESHSTPQVNVWSGYKLFQMIRELGLKIOVMSF
HECGNVGDDVHIOIPWREIGQSPDIYFTISAGRNTTECLTWIDKQRLRGHTA
LEVTFNRSRVERDFEFEEKIPEIEVEGFGCGELRIPSYPAQGWKPIGIGERQC
YKYLMSNLKEAAEVRGHSFWGSPONTETYNSTPHGTGFRDGGDYDSYGRFLNWM
YSRVLIDHGRVLAANLAFEGTCIAKLSGIHWYKTAHAAELTAGYNASNRDGY
GPIAMFKKHDAALNFTCVELRTLDHEDPEALADPEGLVMQVLAANDASIPVASE
NALFCYREGYNKILENAKPLTDPGRHLSCTYLRNLPTLMESQNFKEFERFLKRMH
GEAVPDLGLAPGTOETNPE"
complement(13027..14791)
/gene="F6N23.21"
complement(join(13027..13335,13480..13615,13712..13840,
13943..14001,14087..14791))
/gene="F6N23.21"
/codon_start=1
/evidence=not experimental
/protein_id="AAC13622.1"
/db_xref="GI:3047111"
/translation="MSILCCVPVLECYVLCGCTHNLWKKLYSAGHESNWLGSDE
PEFPRICRILAVYEENHDPWAPDGYGIDFNHVLKDYDQTEGRVTFYMIYLD
HENGDVLAIRGLNAKECDYAVLNLKQTFDGYVHNGLLKAAMVFEHVVYL
RELLEANSVSLTFVGLSGVSLVLEVIQNRVRLGNIRKIRCAIPAPRQMS
LHLAVTVADVNSVLQDDLPRTTALENVKSIICLCLCLCTCKDTFTFEBEKL
KDARLAPGELHYIVRKLRGRYPVTVAVPDGREGQIVLSNATADHAITWI
BRESQRALDKIVROKSDHDEIRAAIMKAASLNIPMSFSPSYGTFTDTEGESSA
GSGMEGSPSGMSGMRKRWQDFIDCHFPVNDNSEHMIKKNQESQALLOQET"
15539..16687
/gene="F6N23.2"
join(15539..15554,15672..15764,15953..16026,16133..16208,
16307..16353,16466..16687)
/gene="F6N23.2"
/Note="Similar to protein kinase; coded for by A. thaliana
cDNA H36947; coded for by A. thaliana cDNA H37158"
/codon_start=1
/protein_id="AAC13612.1"
/db_xref="GI:3047101"
/translation="WNTESVVEFLGNVTLQRLPSSSLKEISEVVFYKGVDRGDYVVR
ENQVGVYFELQGVRAQVLRAGENYQEPFLKRYDFPGHIGEDVYSADVAYTE
VLCLLMSDRHALLKESVSDSRCLVEDILYLEPLDVRVSPNCPSPSSISVLMIK
VLTTHFYHHSLESEL"
16976..19151
/gene="F6N23.3"
join(16976..17111,17262..17415,17785..17889,18316..18417,
18618..18666,19056..19151)
/gene="F6N23.3"
/codon_start=1
/evidence=not experimental
/protein_id="AAC13635.1"
/db_xref="GI:3047124"
/translation="WEIYSNHCDTVYASOALAAAKTVETMKVINFHCYFLLVGDGI
NIPILYDNRLDGNNFATRSVDAROKGTITFLFASFOVSCVPIRIHVESLDYRNO
VATISVFPFIDIRFCENRSTKQNSPPRCVAFASDLIPATISLNPHERGMSVAA
LSLDSHWMFHRPVRADDWLLFVIVSPATSRGFA'GKMFNRRKGEVRCCIEE"
complement(20843..21366)
/gene="F6N23.20"
complement(join(20843..20944,21002..21056,21103..21155,
21292..21366))
/gene="F6N23.20"
/codon_start=1
/evidence=not experimental
/protein_id="AAC13621.1"
/db_xref="GI:3047110"
/translation="MWATAVASVFLLTGTDYGPENVLDPYLSSSMILLVLEVLIRMF
SYWVWLHLFMLEHVIHWIKKILAAQDSVQKFI'GPKGEPKGPSTAK"
complement(22410..24753)

```

Db 1297 ATTAAGATGAGCGGAGACGGCTGCTGTGCTTGGAGGAAAGGTTGATGACTGAAGA 1356  
 Qy 261 GlyLeuAlaAlaGluLysMetAspAlaArgGlyLeuLeuValAlaCysPhe 280  
 Db 1357 GGATTAGCTGGCTGAGAAATGGATGCAAGGGTTGCTTTACTAGTCTCTGTTT 1416  
 Qy 281 GlyValProSerPheArgSerThrAspLeuLeuAspLeuLeuArgMetSerGlySer 300  
 Db 1417 GGTGTTCTCTCAAACTTTAGAGATGACAGATTTCTGATGATTTGATAGGATGAGTGGTGG 1476  
 Qy 301 AsnGluLeuAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetValSer 318  
 Db 1477 AATGAGATGCGCGTGTCTTCAAGCGGTACAGTTTCTTGCTCCTATGCTCTCAGGTACC 1536  
 Qy 318 ----- 318  
 Db 1537 ATATCTGTTCTACTCGGTGAATTTCAATGGAAGGTGGTTCCTTTGTTGACATCATC 1596  
 Qy 318 ----- 318  
 Db 1597 GACCAACATCAAGTTCCATCTTTGTTTTCGATAAGCTTGATGGTATATAAACTAGGAGAC 1656  
 Qy 318 ----- 318  
 Db 1657 ACATCAAAATTTAGAGTGCAATGACTGATGTAGGCCAAATCTTAGTGAATAATTAATCG 1716  
 Qy 318 ----- 318  
 Db 1717 GAAAGAACTTGGAACTCTCAACCATAGTTTTCGTACGAAATTTGTTGCTTGTGACAGACCA 1776  
 Qy 318 ----- 318  
 Db 1777 AATGATAGGCTATTGCTTGAAATAGTGTTCCTTGTGTTTCCAAATTTGGAAGTTAAAA 1836  
 Qy 318 ----- 318  
 Db 1837 TCATATGACTAGCTGTGGTACTAATTAAGCTTAAGCAATGCCAACTCTAAGAAGTGG 1896  
 Qy 319 ----- GlyLeuValGluSerSerIleLysArgGlyMet 329  
 Db 1897 TACTTACACAATATTCATTGTTGGTCATGATGATAGTTGAATCAAGTATCAAGCGTGAATG 1956  
 Qy 330 HisLeuAlaLeuGluMetValThrPheGlyMetGluAspLysPheSerAlaAla 349  
 Db 1957 CATATTGAGCTCTTGAGATGGTTTATACCTTTGGCATGGAGATAAGTTTTCAGTGTCT 2016  
 Qy 350 LeuValLeuThrSerPheLeuLysMetSerLysGluSerPheGluArgAlaLysArgLys 369  
 Db 2017 CTAGTTCTTACTTCACTTCTTAAAGATGAGCAAGGAGTCAATTTGAGAGGCGCAAAACGGAA 2076  
 Qy 370 AlaGlnSerProLeuAlaPheLys----- 377  
 Db 2077 GCCAGTACCGCTGGCATTTGT-ATGAACCTTCCGTTGCACATTATGACTCTTATGA 2135  
 Qy 378 -----GluAla 379  
 Db 2136 ACTCTTTATCATCTGAGTCTGACCATTTATATTTATTTCTCAACAGCAAGAGCG 2195  
 Qy 380 AlaThrLysGlnLeuAlaValLeuSerSerValMetGlnCysMetGluThrHisLysLeu 399  
 Db 2196 GCTCAAGACGCTAGCTGTGTTATCATCATGATGATGATGATGATGATGATGATGATGATGAT 2255  
 Qy 400 AspProAlaLysGluLeuProGlyTrpGlnIleLysGluGlnIleValSerLeuGluLys 419  
 Db 2256 GATCCTCGGAAGAACTACCAAGATGTCAGATCAAGAGCAAAATTTGATGTTGAGAA 2315  
 Qy 420 AspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSerLeuLeuMetGlu 439  
 Db 2316 GACACTCTTACGTCGACAAAGAGATGGAAGAGAAAGCAAGATCTCTCAGTTTAAATGGAG 2375  
 Qy 440 GluAlaAlaLeuAlaLysArgMetValAsnGlnGlnIleLysArgProArgLeuSerPro 459  
 Db 2376 GAAGCCGCACTTGCCAGAGAGATGATATAACACAGATAAAAGTCCAGGTTGTACCC 2435

Qy 460 MetGluMetProProValThrSerSerSerTyrosProIleTyArgAspArgSerPhe 479  
 Db 2436 ATGGAATGCCACAGTAACTTCCTTCATCGTATTTCTCTATCTACCGGTAGTAGAGCTTT 2495  
 Qy 480 ProSerGlnArgAspAspGlnAspGluLeuSerAlaLeuValSerSerTyrosGly 499  
 Db 2496 CCTAGTCAAAGAGACGATGACCAAGATGAATATCATCAGCTCTTGTGAGTAGTACTCTCGGC 2555  
 Qy 500 ProSerThrSerPheProHisArgSerArgArgSerProGluTyMetValProLeuPro 519  
 Db 2556 CCGTCAACATCTTTCTCTCATCGTCAAGAAGATCCCCGAATATATGTTTCCACTTCCA 2615  
 Qy 520 HisGlyGlyLeuGlyArgSerValTyAlaTyGluHisLeuAlaProAsnSerTySer 539  
 Db 2616 CATGGTGGTTAGGAAGAGTATATGCATATGAACATCTGGCCCCAAATTCATCTCT 2675  
 Qy 540 ProGlyHisGlyHisArgLeuHisArgGlnTySerProSerLeuValHisGlyGluArg 559  
 Db 2676 CCAGGTCAAGGACATAGACTTCTCCATTCATGGACAACACAGTTACCATATGTTATACAA 2735  
 Qy 560 HisProLeuGlnTySerProProIleHisGlyGlnGlnLeuProTyGlyIleGln 579  
 Db 2736 CATCCACTACGACTCTCTCTCCATTCATGGACAACACAGTTACCATATGTTATACAA 2795  
 Qy 580 ArgValTyArgHisSerProSerGluGluArgTyLeuGlyLeuSerAsnGlnArgSer 599  
 Db 2796 AGGGTTTACAGACATTCACCATCTGAAGAAGATATTTGGGTTTATCCATCAAGGTCT 2855  
 Qy 600 ProArgSerAsnSerSerLeuAspProLys 609  
 Db 2856 CTTCCGAGTAACTCATCATAGACCCCA 2885

RESULT 10  
 F6N23  
 LOCUS 91849 bp DNA linear PLN 12-NOV-1999  
 DEFINITION Arabidopsis thaliana BAC F6N23.  
 ACCESSION AF058919  
 VERSION AF058919.2 GI:6382042  
 KEYWORDS  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
 REFERENCE 1 (bases 1 to 91849)  
 Geisel, C.  
 The sequence of A. thaliana F6N23  
 Unpublished  
 REFERENCE 2 (bases 1 to 91849)  
 Washington University Genome Sequencing Center.  
 The A. thaliana Genome Sequencing Project  
 Unpublished  
 REFERENCE 3 (bases 1 to 91849)  
 Waterston, R.  
 Direct Submission  
 Submitted (10-APR-1998) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 REFERENCE 4 (bases 1 to 91849)  
 Waterston, R.  
 Direct Submission  
 Submitted (12-NOV-1999) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 On Nov 12, 1999 this sequence version replaced gi:3047100.  
 COMMENT Submitted by:  
 Genome Sequencing Center  
 Department of Genetics, Washington University,  
 St. Louis, MO 63108, USA  
 e-mail: rwaterston@wustl.edu

MAPPING: Clones were assigned to the YAC map by hybridization by  
 M. Lodhi, Cold Spring Harbor Laboratories, and fingerprinted

Db 2086 ACTCTTTATCATCATCTGAGTCTGACCATGATGATATATTTATTTCTCAACAGAAAGAGCG 2145

QY 380 AlaThrLysGluLeuAlaValLysSerValMetGlnCysMetGluThrHisLysLeu 399

Db 2146 GCTACAAAGCAGCTAGCTGTGTATCATCAGTATGTCAGTGTATGGAGACTCAAGTTA 2205

QY 400 AspProAlaLysGluLeuProGlyTrpGlnIleLysGluGlnIleValSerLeuGluLys 419

Db 2206 GATCTCTGGAAAGAACTACCAAGATGGCAGATCAAGAGCAAAATGTTAGCTTGAGAAA 2265

QY 420 AspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSerLeuSerLeuMetGlu 439

Db 2266 GACACTCTTCAGCTCGACAAAGATGAAGAGAAAGCAAGATCTTCAGTTTAATGGAG 2325

QY 440 GluAlaLeuAlaLysArgMetTyrAsnGlnIleLysArgProArgLeuSerPro 459

Db 2326 GAAGCGCACTTGGCAAGAGATGATATCAACACAGATAAAACGTCCAAGTTGTCAACC 2385

QY 460 MetGluMetProProValThrSerSerTyrSerProLietTyrArgAspArgSerPhe 479

Db 2386 ATGGAAATGCCACCAAGTAACTTCITCATCGTATCTCTATCTACCGTGATAGAAGCTTT 2445

QY 480 ProSerGlnArgAspAspGlnAspGluIleSerAlaLeuValSerSerTyrLeuGly 499

Db 2446 CTAGTCAAGAGACGATGACCAAGATGAATATATCAGCTCTGTGAGTAGTTACTCGGC 2505

QY 500 ProSerThrSerPheProHisArgSerArgArgSerProGluTyrMetValProLeuPro 519

Db 2506 CGGTCAACATCTTTTCCTCATCGCTCAAGAGATCCCGGAATATATGGTTCCACTTCCA 2565

QY 520 HisGlyGlyLeuGlyArgSerValTyrAlaTyrGluHisLeuAlaProAsnSerTyrSer 539

Db 2566 CATGGTGGTTAGGAGAGGTATATGCATATGACATCTGGCCCCAATTCATCTCT 2625

QY 540 ProGlyHisGlyHisArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGlnArg 559

Db 2626 CCAGGTTCACGACATAGACTTCATCGACAGTACTCTCGTCTTGGTTCCGGACAGAGA 2685

QY 560 HisProLeuGlnTyrSerProPheHisGlyGlnGlnGlnLeuProTyrGlyIleGln 579

Db 2686 CATCCACATACAGTACTCTCCCTCAATTCATGACAAACAGAGTTACCATATGGTATACAA 2745

QY 580 ArgValTyrArgHisSerProSerGluGluArgTyrLeuGlyLeuSerAsnGlnArgSer 599

Db 2746 AGGGTTTACAGACATTCACCATCTGACAAAGAGATTTTGGTTTATCCATCAAGGTCT 2805

QY 600 ProArgSerAsnSerSerLeuAspProLys 609

Db 2806 CCTCGAGTAACATCATATTAGACCCCAAA 2835

RESULT 9

AV198404 3758 bp DNA linear PLN 12-JUN-2003

LOCUS Arabidopsis thaliana ecotype Cvi FRIGIDA-like protein gene,

DEFINITION complete sequence.

ACCESSION AV198404

VERSION AV198404.1 GI:31558916

KEYWORDS Arabidopsis thaliana (chale cress)

SOURCE Arabidopsis thaliana

ORGANISM Arabidopsis thaliana

REFERENCE 1 (bases 1 to 3758)

AUTHORS Gazzani, S., Gendall, A.R., Lister, C. and Dean, C.

TITLE Analysis of the molecular basis of flowering time variation in Arabidopsis accessions

JOURNAL Plant Physiol. 132 (2), 1107-1114 (2003)

REFERENCE 2 (bases 1 to 3758)

AUTHORS Gazzani, S. and Dean, C.

TITLE Direct Submission

JOURNAL Submitted (17-DEC-2002) Cell and Developmental Biology, John Innes Center, Colney Lane, Norwich NR4 7UH, UK

FEATURES

Source 1..3758

Location/Qualifiers

/organism="Arabidopsis thaliana"

/mol\_type="genomic DNA"

/db\_xref="taxon:3702"

/chromosome="4"

/map="between mi51 and mi122"

/note="ecotype: Cvi"

misc\_feature 577..2888

/note="similar to FRIGIDA (FRI)"

BASE COUNT 1178 a 714 c 774 g 1092 t

ORIGIN

Alignment Scores:

Pred. No.: 7,09e-185 Length: 3758

Score: 2950.00 Matches: 604

Percent Similarity: 78.57% Conservative: 1

Best Local Similarity: 78.44% Mismatches: 4

Query Match: 95.04% Indels: 162

DB: 8 Gaps: 2

US-09-890-475-1 (1-609) x AV198404 (1-3758)

QY 1 MetSerAsnTyrProProThrValAlaAlaGlnProThrThrThrAlaAsnProLeuLeu 20

Db 577 ATGTCCAATTATCACCAGCGGTGGCGCGCAACCAACAGCGCGAATCCACTGCTG 636

QY 21 GlnArgHisGlnSerGluGlnArgArgGluLeuProLysIleValGluThrGluSer 40

Db 637 CAGCGACATCAATCTGAACAGCGAGAGAGATTACCGAGATTGTCGAACAGAGTCT 696

QY 41 ThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLysSerIleAsp 60

Db 697 ACAAGTAGGCATTTACGATCGGTCAATCTTAAGCAGCTCAATTTTGAATCCATAGAC 756

QY 61 GluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAspLeuGln 80

Db 757 GAATTAGCTGGTTTTCAGTTGAGTGGAGAAACATTCAATGCCAATTCGATGATCTTCAG 816

QY 81 LysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsnGlyValVal 100

Db 817 AAGCACATCGAGTCAATCGAAACCGCAATTCATTCGAGAGTAAACGGCGTTGTC 876

QY 101 LeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSerProProArgAsnVal 120

Db 877 CTCGCCGCGCGAGACAAATATTTCCATCAGCCGATGTTATCGCTCCGCGAGCAATGTA 936

QY 121 SerValGluThrThrValThrValSerGlnProSerGlnGluIleValProGluThrSer 140

Db 937 TCTGTAGAAACCCAGCTCACTGTGAGCCAAACCGCTCTCAGGAGATTGTACCGGAGACGTG 996

QY 141 AsnLysProGluGlyGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyr 160

Db 997 AATAAACCGAGGGGGGACCGTATGTGTGAGTTGATGTAGCAAGGTCTGCGTAATAATAC 1056

QY 161 IleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGluGluIleProSerAlaLeuLys 180

Db 1057 ATATACGCAATATCTCTGAACAGCTAAGTTAATGGAAGAGATTCTTCAGCTTTGAAA 1116

QY 181 LeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPheTyrLeuGlnGly 200

Db 1117 TTGCCAAGGAGCGAGAGTTGATTGAGTTGATTGAGTTGATTGAGTTGAGTTGAGTTGAG 1176

QY 201 ArgArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuIle 220

Db 1177 CGTAGACATTTACTAAGAGTGCCTATGAGTCTTCGAGACAAAGTTTCGCTTCTTATA 1236

QY 221 LeuGluSerPheLeuLeuMetProAspArgGlyLysGlyLysValLysIleGluSerTrp 240

Db 1237 CTGGAGTCTTTCTTCTAATGCTGATCGTGGTTAAGGGAAGGTGAAGATTGAGAGTTGG 1296

QY 241 IleLysAspGluAlaGluThrAlaAlaValAlaTyrArgLysArgLeuMetThrGluGly 260



```

Db      1114 TTGGCCAGGACCCAGCGAAGTTTATTTGGATTGTATGGCAAGTTTTTACTTACAGGG 1173
QY      201 ArgArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuIle 220
Db      1174 CGTAGAGCATTTACTAAAGAGTGGCTATGAGCTCTGCGACAGATTTTCGCTTCATTATA 1233
QY      221 LeuGluSerPheLeuLeuMetProAspArgGlyLysGlyValLysIleGluSerTrp 240
Db      1234 CTGGAGTCTTTTCTTCTATGCTGATCGTGGTAAGGAAGGTGAAGATGGAGATGG 1293
QY      241 IleLysAspGluAlaGluThrAlaAlaValAlaTrpArgLysArgLeuMetThrGluGly 260
Db      1294 ATTTAAAGATGAGCGCGGACGCGTCTGCTGCTTGGAGGAAAGGTTGATGATCGAGGA 1353
QY      261 GlyLeuAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuValAlaCysPhe 280
Db      1354 GGATTAGCTGGCGCTGGAAATGGATGCAAGGGGTTTGCTTTTACTAGTTGCTTGTGTTT 1413
QY      281 GlyValProSerAsnPheArgSerThrAspLeuLeuAspLeuIleArgMetSerGlySer 300
Db      1414 GGTGTTCTTCAAACCTTTAGCGATACAGATTTGCTGGATTTGTAAAGATGAGTGGTTCG 1473
QY      301 AsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetValSer----- 318
Db      1474 AATGAGATTGCGGCTGTTTGAAGCGGTACAGTTTCTGTGCTATGTTCTCAGGTACC 1533
QY      318 ----- 318
Db      1534 ATATTCTGTTCTACCTCGGTGANTTTCATTCCAAAGTGGTTCCTTTTGTGTGACATC 1593
QY      318 ----- 318
Db      1594 GACCAACATCAAGTTCCATCTTTGTTTTCGATAAGCTTGATGTTATAAACTAGGAGAGC 1653
QY      318 ----- 318
Db      1654 ACATCAATATTTAGAGTGCAATGACTGATTGAGCCAAATCCTAGCTAGAAATTAATCTG 1713
QY      318 ----- 318
Db      1714 GAAAGAACTTGGAACTCTCAACCATAGGTTTGGTACGAAATTTGCTGTTGTCAGAACCA 1773
QY      318 ----- 318
Db      1774 AATGATAGCTATTGCGCTTGAATAGTGTTCCTTGTGTTTCCAAATATCGAAGTAAAA 1833
QY      318 ----- 318
Db      1834 TCATATGACTTAGCTGTGGATACATACTAATTAAGCTTAAGCAATGCCAATCTAAGAAGTGG 1893
QY      319 ----- GlyIleValGluSerSerIleLysArgGlyMet 329
Db      1894 TACTTACACAAATTTCTATTGTTGTCATAGGTATAGTTGAATCAAGTATCAAGCGTGAATG 1953
QY      330 HisIleGluAlaLeuGluMetValThrPheGlyMetGluLeuAspLysPheSerAlaAla 349
Db      1954 CATATTGAAGCTCTTGAGATGGTTTATACCTTTGCGATGGAGGATAAGTTTTCAGTGCT 2013
QY      350 LeuValLeuThrSerPheLeuLysMetSerLysGluSerPheGluArgAlaLysArgLys 369
Db      2014 CTAGTTCTTAACCTTCACTTAAAGATGAGCAAGGAGTCATTTGAGAGGGCAAAACGGAAA 2073
QY      370 AlaGlnSerProLeuAlaPheLys----- 377
Db      2074 GCCCAGTACCCTGGCAATTTGT-ATGAACCTTCCTCGCATATATGATGACCTTTATGA 2132
QY      378 -----GluAla 379
Db      2133 ACTCTTTATCATCTGAGTCTGACCATTCATATATTTTCTCAACAGAAAGAGCG 2192
QY      380 AlaThrLysGlnLeuAlaValLeuSerSerValMetGlnCysMetGluThrHisLysLeu 399

```

```

Db      2193 GCTACAAAGAGCAGCTAGCTGTGTTTATCATCATGATTATGCAAGTGTATGAGACTCACAAGTTA 2252
QY      400 AspProAlaLysGluLeuProGlyTrpGlnIleLysGluGlnIleValSerLeuGluLys 419
Db      2253 GATCTCGAAAGAACTACCAAGGATGCGAGATCAAAAGAGCAAAATTTAGCTTGGAGAAA 2312
QY      420 AspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSerLeuSerLeuMetGlu 439
Db      2313 GACACTCTTACGCTCGAAGAGATGGAAGAGAAAGCAAGATCTCTCAGTTTAAATGGAG 2372
QY      440 GluAlaAlaLeuAlaLysArgMetTyrAsnGlnGlnIleLysArgProArgLeuSerPro 459
Db      2373 GAAGCGCACCTTGCAGAGAGATGATATAACCAACAGATAAAACGTCCAAAGTTGTCAACC 2432
QY      460 MetGluMetProProValThrSerSerTyrSerProIleTyrArgAspArgSerPhe 479
Db      2433 ATGGAATGTCACCAAGTAACCTTCTCATCGTATTTCTCTATCTACCGTGATAGAAGCTTT 2492
QY      480 ProSerGlnArgAspAspGlnAspGluIleSerAlaLeuValSerSerTyrLeuGly 499
Db      2493 CCTAGTCAAAGAGACGATGACCAAGATGAATATCAGCTCTTGAGTAGTTACCTCGGC 2552
QY      500 ProSerThrSerPheProHisArgSerArgArgSerProGluTyrMetValProLeuPro 519
Db      2553 CCGTCAACATCTTTTCTCATCGCTCAAGAAGATCCCGGAATATATGTTCCACTTCCA 2612
QY      520 HisGlyGlyLeuGlyArgSerValTyrAlaTyrGluHisLeuAlaProAsnSerTyrSer 539
Db      2613 CATGGTGGTTAGGAAGAGTGTATATGATATGAACATCTGGCCCCCAATTCATACTCT 2672
QY      540 ProGlyHisGlyHisArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGlnArg 559
Db      2673 CCAGTCCAGGACATAGACTTCATCGACAGTACTCTCCGTCTTTGGTTCCAGGACAGAGA 2732
QY      560 HisProLeuGlnTyrSerProIleHisGlyGlnGlnLeuProTyrGlyIleGln 579
Db      2733 CATCACTACAGTACTCTCTCCANTCATGGACACACAGATTACCATATGTTATACAA 2792
QY      580 ArgValTyrArgHisSerProSerGluGluArgTyrLeuGlyLeuSerAsnGlnArgSer 599
Db      2793 AGGGTTTACAGACATTCACCATCTGAAGAAGATATTGGGTTTATCCAAATCAAAGGTCT 2852
QY      600 ProArgSerAsnSerSerLeuAspProLys 609
Db      2853 CCTCGAGTAACATCATTAGACCCCAAA 2882

RESULT 8
AY198405
LOCUS   Arabidopsis thaliana ecotype Wil-2 FRIGIDA (FRI) gene, complete
DEFINITION
VERSION AY198405.1 GI:31558917
KEYWORDS
SOURCE  Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE
  1 (bases 1 to 3301)
AUTHORS   Gazzani, S., Gendall, A.R., Lister, C. and Dean, C.
TITLE     Analysis of the molecular basis of flowering time variation in
Arabidopsis accessions
JOURNAL   Plant Physiol. 132 (2), 1107-1114 (2003)
REFERENCE
  2 (bases 1 to 3301)
AUTHORS   Gazzani, S. and Dean, C.
TITLE     Direct Submission
JOURNAL   Submitted (17-DEC-2002) Cell and Developmental Biology, John Innes
Center, Colney Lane, Norwich NR4 7UH, UK
FEATURES
    source
        1..3301
            /organism="Arabidopsis thaliana"
            /mol_type="genomic DNA"

```

```

Db      2184  GATCCTCGAAGAAGAACTACACGAGGATGCCAGATCAAGAGCAAAATTGTTAGCTTGAGAGAA 2243
Qy      420  AspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSerLeuSerLeuMetGlu 439
Db      2244  GACACTCTTCAGCTCGACAAAGAGATGGAAGAGAAAGCAAGATCTCTCAGTTTAAATGGAG 2303
Qy      440  GluAlaAlaLeuAlaLysArgMetTyrAsnGlnGlnLysArgProArgLeuSerPro 459
Db      2304  GAAGCCGACCTTGCCAAAGAAATGTATACCAACAGATAAACGTCACAGGTGTCACCC 2363
Qy      460  MetGluMetProValThrSerSerTyrSerProIleTyrArgAspArgSerPhe 479
Db      2364  ATGGAATGCCACAGTAATCTTCTTCATCGTATTCTCTCATCTACCGTGATAGAGCTTT 2423
Qy      480  ProSerGlnArgAspAspGlnAspGluLysSerAlaLeuValSerSerTyrLeuGly 499
Db      2424  CCTAGTCAAGAGACGATGACCAAGATGAATATACAGCTCTTGAGTAGTTTACCTCGC 2483
Qy      500  ProSerThrSerPheProHisArgSerArgArgSerProGluTyrMetValProLeuPro 519
Db      2484  CCGTCACATCTTTCTCATCGCTCAAGAGATCCCGGAATATATGTTTCCACTTCCA 2543
Qy      520  HisGlyGlyLeuGlyArgSerValTyrAlaTyrGluHisLeuAlaProAsnSerTyrSer 539
Db      2544  CATGTCGGGTAGGAAGAGTGATATGCATATGAACATCTGCGCCCAAAATTCATACTCT 2603
Qy      540  ProGlyHisGlyHisArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGlnArg 559
Db      2604  CCAGTTCAGGACATAGACTTCATCGACGATCTCTCCGCTCTTGGTTCACGACAGAGA 2663
Qy      560  HisProLeuGlnTyrSerProIleHisGlyGlnGlnLeuProTyrGlyLeuGln 579
Db      2664  CATCCACTACAGTACTCTCTCCAAATTCATGGAACAACAGATTACCATATGATATACAA 2723
Qy      580  ArgValTyrArgHisSerProSerGluGluArgTyrLeuGlyLeuSerAsnGlnArgSer 599
Db      2724  AGGTTTACAGACATTCACCATCTGAAGAAGATATTTGGTTTATTCCAATCAAAAGGCT 2783
Qy      600  ProArgSerAsnSerSerLeuAspProLys 609
Db      2784  CCTCGCAGTAACATCATCTAGACCCCAAA 2813

RESULT 7
LOCUS   AY198401
DEFINITION   Arabidopsis thaliana ecotype Shakhdara FRIGIDA (FRI) gene, complete cds.
ACCESSION   AY198401
VERSION     AY198401.1 GI:31558910
KEYWORDS   Arabidopsis thaliana (thale cress)
SOURCE     Arabidopsis thaliana
ORGANISM   Arabidopsis thaliana
REFERENCE  1  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS    Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
TITLE       rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
JOURNAL    1  (bases 1 to 3739)
AUTHORS    Gazzani, S., Gendall, A.R., Lister, C. and Dean, C.
TITLE       Analysis of the molecular basis of flowering time variation in
JOURNAL    Arabidopsis accessions
REFERENCE  2  Plant Physiol. 132 (2), 1107-1114 (2003)
AUTHORS    Lister, C. and Dean, C.
TITLE       Direct Submission
JOURNAL    Submitted (17-DEC-2002) Cell and Developmental Biology, John Innes
FEATURES   Center, Colney Lane, Norwich NR4 7UH, UK
            Location/Qualifiers
            1..3739
            /organism="Arabidopsis thaliana"
            /mol_type="genomic DNA"
            /db_xref="taxon:3702"
            /chromosome="4"
            /map="between ml51 and ml122"
            /note="ecotype: Shakhdara"

```

```

gene
mRNA
CDS
BASE COUNT      1172 a      711 c      767 g      1089 t
ORIGIN
Alignment Scores:
Pred. No.:      1 59e-186      Length:      3739
Score:          2975.00      Matches:      607
Percent Similarity: 78.83%      Conservative: 0
Best Local Similarity: 78.83%      Mismatches: 2
Query Match:      95.84%      Indels:      162
DB:              8      Gaps:      2

US-09-890-475-1 (1-609) x AY198401 (1-3739)
Qy      1  MetSerAsnTyrProProThrValAlaAlaGlnProThrThrThrAlaAsnProLeuLeu 20
Db      574  ATGTCCAATTATTCACCGACGGTGGCGGCAACCCACACGACGCGCAATCCCACTGCTG 633
Qy      21  GlnArgHisGlnSerGluGlnArgArgGluLeuProLysIleValGluThrGluSer 40
Db      634  CAGCGACATCAATCTGAACAGCAGCAGAGAGATTTACCGAGATTTGTCGAACAGAGTCT 693
Qy      41  ThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLysSerIleAsp 60
Db      694  ACAAGTATGACATTCAGATCGGTCAATCTAAGCAGCCTCAAAATTTGAAATTCATAGAC 753
Qy      61  GluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAspAspLeuGln 80
Db      754  GAATTAGTCGCTTTTCAGTTGAGTGGAAACATTTCAAACCGCAATTCGATGATCTTCAG 813
Qy      81  LysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsnGlyValVal 100
Db      814  AAGCACATCGAGTCAATCGAAACCAATTTGATTTCAAATCTCGAGAGTAACGGCGTTGTC 873
Qy      101  LeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSerProProArgAsnAsnVal 120
Db      874  CTGCGCGCGGAGCAATAAATTTCCATCAGCGGATGTTATCGCTCCCGGAAACAATGTA 933
Qy      121  SerValGluThrThrValThrValSerGlnProSerGlnGluIleValProGluThrSer 140
Db      934  TCTGTAGAAACCAACCGTCACTGTGAGCCAAACCGTCTCAGGAGATTGTACCGGAGACGTCG 993
Qy      141  AsnLysProGluGlyGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyr 160
Db      994  AATAAAACCGGAGGCGGACGTATGTGTGAGTTGATGTGTAGCAAAAGGTCTCGGTAATAAC 1053
Qy      161  IleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGluGluIleProSerAlaLeuLys 180
Db      1054  ATATAGCGGAATATCTCTGATCAAGCTAAGTTAATGAGAGAGATTCTCTTACGCTTTGAA 1113
Qy      181  LeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPheTyrLeuGlnGly 200

```

```

/ product="FRIGIDA"
Join(505, .1459,1853, .2025,2115, .2816)
/ gene="PRI"
/ note="flowering gene"
/ codon_start=1
/ product="FRIGIDA"
/ protein_id="AAP49808.1"
/ db_xref="GI:31558913"
translation="MSNYPTVAQPTTANPLLQHQSORRELKPIVETSTSD
ITIGSKPOLIKSIDELAAESVAVEITFRQFDLQKHIESIENAIQKUESGVIA
ARNNPHOMUSPPRNNSVETTVTSQSPQEIIVFETSNKPEGGRMCELMCSKGLRKY
IYANTSDAKLMEETPSALKAKPAKFEVDCIKFYLOQRRAFTKSPMSARQVSL
LLESLFLMDPRGKGKVKIISWIKDEAETAAVAWKRLMTGGLAAAEKMDARGLLIL
VACVFGNFRSTDLDLIRMSGNEIAGALKRSQFLVPMVSGIVESSIKRGWHIHAL
EMVYTFMGEDKFSALVLTSLKMSKESFERAKKAQSLAFKEAATKOLAVLSSVMO
CMETHLPDAPKELPWQIKELIVSLKDTLQDKEMEREKARSLSIMEAALAKRMNQ
QIKRPLSPMPPTSSSYPTFRDRSPFSCRDDDDQDEISALVSSVLSPGSTSPFPHRS
RRSPYVPLPHMGGRSVYAYEHLAPNSYSPGHRLHRQYSPSLVHGHRHLPQYSP
PIHGQOLPYGIQVRYRHSPEERYLGLSNQSRSPNSSLDPK"
BASE COUNT 990 a 638 c 666 g 967 t
ORIGIN

Alignment Scores:
Pred. No.: 1,33e-186 Length: 3261
Score: 2975.00 Matches: 607
Percent Similarity: 78.83% Conservative: 0
Best Local Similarity: 78.83% Mismatches: 2
Query Match: 95.84% Indels: 162
DB: 8 Gaps: 2

US-09-890-475-1 (1-609) x AY198402 (1-3261)

QY 1 MetSerAsnTyrProProThrValAlaAlaGlnProThrThrAlaAsnProLeuLeu 20
Db 505 ATGTCCAATTATCCACGCGGTGCGCGCAACCCACACGCGCGCAATCCACTGCTG 564
QY 21 GlnArgHisGlnSerGluGlnArgArgGlnLeuProLysIleValGluThrGluSer 40
Db 565 CAGCGACATCAATCTGAACAGCGAGAGAGAAATACCGAAGATTGCGAAACAGAGTCT 624
QY 41 ThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLysSerIleAsp 60
Db 625 ACAAGTATGGACATTACGATCGGTCAATCTAAGCAGCGCTCAAAATTTTGAATCCATAGAC 684
QY 61 GluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAspAspLeuGln 80
Db 685 GAATTAGTCTGCGTTTTCAGTTGCAAGTGAACCAATTCATCAACGCCAATTCGATGATCTTCAG 744
QY 81 LysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsnGlyValVal 100
Db 745 AAGCATATCGAGTCAATCGAAAAACCAATTCATTCCAAACTCGAGAGTAACGGCGTTGTC 804
QY 101 LeuAlaAlaArgAsnAsnAspPheHisGlnProMetLeuSerProArgAsnAsnVal 120
Db 805 CTCGCGCGCGCGAAACAATAATTTCCATCAGCCGATGTTATCGCCCTCGCGCGAAACAATGTA 864
QY 121 SerValGluThrThrValThrValSerGlnProSerGlnGluIleValProGluThrSer 140
Db 865 TCTGTAGAAACACCCGTCCTGTAGCCCAACCGTCTCAGGAGATTGACCGGAGCGTCG 924
QY 141 AsnLysProGluGlyGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyr 160
Db 925 AATAAACCGAGGGGGGCGATGTGTGTGAGTTGATGTGTAGCAAGGTCGCGTAAATAC 984
QY 161 IleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGluGluIleProSerAlaLeuLys 180
Db 985 ATATACCGGAATATCTCGATCAAGCTAAGTTAATGAAGAGATTCCTTCAGGTTTGA 1044
QY 181 LeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPheTyrLeuGlnGly 200
Db 1045 TTGGCCAAAGGACCGACGAGATTGTTGATTGATTGATTGATTGATTGATTGATTGATTGATTG 1104
QY 201 ArgArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuIle 220

```

```

Db 1105 CGTAGAGCATTTACTAAAGAGTCGCCTATGAGCTCTCGGAGACAAAGTTTCGCTCTTTATA 1164
QY 221 LeuGluSerPheLeuLeuMetProAspArgGlyLysGlyLysValLysIleGluSerTyr 240
Db 1165 CTGGAGTCTTTCTTCTTAATGCCTGATCGTGGTAAAGGAAGGTGAAGATTGAGAGTTGG 1224
QY 241 IleLysAspGluAlaGluThrAlaAlaValAlaIleThrPArgLysArgLeuMetThrGluGly 260
Db 1225 ATTAAGATGAGGCGGAGACGCGCTGCTGTGCTGGAGGAAAAGTTGATGACTGAAGGA 1284
QY 261 GlyLeuAlaAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuValAlaCysPhe 280
Db 1285 GGAATTAGCTCGGCTGAGAAAATGATGCAAGGGTTGCTTTACTAGTTGCTGTTT 1344
QY 281 GlyValProSerAsnPheArgSerThrAspLeuLeuAspLeuIleArgMetSerGlySer 300
Db 1345 GGTGTCTCTCAAACTTTTAGGAGTACAGATTGCTGATTTGATGAAGGATGAGTGGTTCTG 1404
QY 301 AsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetValSer 318
Db 1405 AATGAGATTGGCGGTGCTTTGAACGGGTACAGTTTCTTCTCCCTATGGTCTCAGGTACC 1464
QY 318 ----- 318
Db 1465 ATATTCTGTTCTCCTCGGTGAATTTCAATTGCAAGGTGCTTCTTTTGTGACATCATC 1524
QY 318 ----- 318
Db 1525 GACCAACATCAAGTTCCATCTTTGTTTCGATAAGCTTGTATGATGATATAAACTAGGAGAGC 1584
QY 318 ----- 318
Db 1585 ACATCAAATATTTAGAGTGCAATGACTGATTGAGCCAAATCCTTAGCTAGAAAATTAATCTG 1644
QY 318 ----- 318
Db 1645 GAAAGAACTTGGAACTCTCAACCATAGGTTTGGTAGAAAATTTGCTTGTGTCAGAACCA 1704
QY 318 ----- 318
Db 1705 AATGATAGGCTATTGCGCTTGAAATAGTGTTCCTTGTGTGTTTCCCAATATTGGAAGTTAAA 1764
QY 318 ----- 318
Db 1765 TCATATGACTTAGCTGTGATGATACTAATAAGCTTAAGCAATGCCAACTCTAAGAAAGTGG 1824
QY 319 -----GlyIleValGluSerSerIleLysArgGlyMet 329
Db 1825 TACTTACACAATATTCATTGGTCATAGGTATAGTTGAATCAAGTATCAAGCGTGAATG 1884
QY 330 HisIleGluAlaLeuGluMetValTyrThrPheGlyMetGluAspLysPheSerAlaAla 349
Db 1885 CATATTGAAGCTCTTGAGATGGTTTATACCTTTGGCATGGAGGATAAGTTTTCAGCTGCT 1944
QY 350 LeuValLeuThrSerPheLeuLysMetSerLysGluSerPheGluArgAlaLysArgLys 369
Db 1945 CTAGTTCTTAATCTTCAATCTTAAAGATGAGCAGAGGTCAATTTGAGAGGGGCAAAACGAAA 2004
QY 370 AlaGlnSerProLeuAlaPheLys----- 377
Db 2005 GCCCAGTCACCGCTGGCATTTGT-ATGAACCTTCCCTTGCACATTTATGTACCTTTATGA 2063
QY 378 -----GluAla 379
Db 2064 ACTCTTTATCATCATCTGAGTCTGACCATTTGATATATTTATTTCTCAACAGAAAGAGCG 2123
QY 380 AlaThrLysGlnLeuAlaValLeuSerSerValMetGlnCysMetGluThrHisLysLeu 399
Db 2124 GCTACAAAGCAGCTAGCTGTGTTATCATCAGTTATCGAGTGTATGAGACTACAAAGTTA 2183
QY 400 AspProAlaLysGluLeuProGlyTyrGlnIleLysGluGlnIleValSerLeuGluLys 419

```

QY 221 LeuGluSerPheLeuLeuMetProAspArgGlyValylsValylsleuSerTirp 240  
 Db 1054 CTGGAGTCCTTTCTTCTAATGCTGATCGTGGTAAAGGAGGTGAGATTGAGAGTTGG 1113  
 QY 241 IleLysAspGluAlaGluThrAlaAlaValAlaTirpArgLysArgLeuMetThrGluGly 260  
 Db 1114 ATTAAAGATGAGCGGAGAGCGGTGCTGTTGCTTGGAGGAAAGGTGATGACTGAAGA 1173  
 QY 261 GlyLeuAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuValAlaCysPhe 280  
 Db 1174 GGATTAGCTGGGCTGAGAAATGGATGCAAGGGTTGCTTTACTAGTTGCTGTTT 1233  
 QY 281 GlyValProSerAenPheArgSerThrAspLeuLeuAspLeuIleArgMetSerGlySer 300  
 Db 1234 GGTGTCCTTCAAACTTTAGGAGTACAGATTGCTGGATTGATAGGAGTGGTGGTTCG 1293  
 QY 301 AsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetValSer 318  
 Db 1294 AATGAGATTGCGGTGCTTTGAAGCGGTACAGTTTCTTGCCCTATGCTCTCAGGTACC 1353  
 QY 318 ----- 318  
 Db 1354 ATATTCTGTTCTCACTCGGTGAATTCATTGCAAAAGGTGTTCTCTTTGTTGACATCATC 1413  
 QY 318 ----- 318  
 Db 1414 GACCAACATCAAGTTCATCTTTGTTTTCGATAAGCTTGATGGTATAAATAGGAGGC 1473  
 QY 318 ----- 318  
 Db 1474 ACATCAAAATATTTAGAGTGCAATGACTGATTTGAGCCAAATCCTAGCTAGAAATTAATCTG 1533  
 QY 318 ----- 318  
 Db 1534 GAAAGAACTTGGAACTCTCAACATAGGTTTGGTAGCAAAATGTTGCTTGTCAGAAACCA 1593  
 QY 318 ----- 318  
 Db 1594 AATGATAGGCTATTGCTTGAATAGTGTTCCTTGTGTTTCCAAATATTGGAAGTTAAAA 1653  
 QY 318 ----- 318  
 Db 1654 TCATATGACTTAGCTTGGTATGTAATAAGCTTAAGCAATGCCAATCTTAAGAAGTGG 1713  
 QY 319 ----- GlyIleValGluSerSerIleLysArgGlyMet 329  
 Db 1714 TACTTACAAATATTCATTGGTCATAGGTATAGTTGAATCAAGTATCAAGCGTGAATG 1773  
 QY 330 HisIleGluAlaLeuGluMetValTyrThrPheGlyMetGluAspLysPheSerAlaAla 349  
 Db 1774 CATATTGAAGCTCTTGAGATGGTTTATACCTTTGGCATGGAGGATAGTTTTCAGCTGCT 1833  
 QY 350 LeuValLeuThrSerPheLeuLysMetSerLysGluSerPheGluArgAlaLysArgLys 369  
 Db 1834 CTAGTTCTAATCTATCTTAAAGATGAGCAAGAGTCAATTTGAGAGGGCAAAACGGAAA 1893  
 QY 370 AlaGlnSerProLeuAlaPheLys ----- 377  
 Db 1894 GCCCAGTCACCGCTGCCATTGTT-ATGAACCTTCTCCCTTGCACATTATGTACCTTTATGA 1952  
 QY 378 ----- GluAla 379  
 Db 1953 ACTCTTTATCATCTCTGAGTCTGACCATTTGATATATTTTCTTCAACGAAGAGCG 2012  
 QY 380 AlaThrLysGlnLeuAlaValLeuSerSerValMetGlnCysMetGluThrHisLysLeu 399  
 Db 2013 GCTACAAAGCAGCTAGCTGTGTATCATCAGTTATGATGATGAGTATGAGAGTCAAGTTA 2072  
 QY 400 AspProAlaLysGluLeuProGlyTirpGlnIleLysGluGlnIleValSerLeuGluLys 419  
 Db 2073 GATCCTCGGAAAGAACTACCGAGATGGCAGATCAAGAGCAAAATTTGTAGCTTGGAGAAA 2132  
 QY 420 AspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSerLeuMetGlu 439

Db 2133 GACACTCTTCACTCGACAAAGAGATGGAAGAGAAAGCAAGATCTCTCAGTTTAAATGAG 2192  
 QY 440 GluAlaAlaLeuAlaLysArgMetTyrAsnGlnGlnIleLysArgProArgLeuSerPro 459  
 Db 2193 GAAGCGGCACCTTCCCAAGAGATGTATTAACCAACAGATAAAACGTCCTCAAGGTTCTCACC 2252  
 QY 460 MetGluMetProProValThrSerSerSerTyrSerProIleTyrArgAspArgSerPhe 479  
 Db 2253 ATGAAATCCCAACAGTAACTTCTTCATCGTATTTCTCTATCTACCGTGATAGAGGTTT 2312  
 QY 480 ProSerGlnArgAspAspGlnAspGluIleSerAlaLeuValSerSerTyrLeuGly 499  
 Db 2313 CCTAGTCAAGAGACGATGACCAAGATGAATATCAGCTCTTGTGAGTAGTTACTCGGC 2372  
 QY 500 ProSerThrSerPheProHisArgSerArgSerProGluTyrMetValProLeuPro 519  
 Db 2373 CCGTCAACATCTTTTCTCATCGCTCAAGAAGATCCCGGAATATATGTTTCCACTTCCA 2432  
 QY 520 HisGlyGlyLeuGlyArgSerValTyrAlaTyrGluHisLeuAlaProAsnSerTyrSer 539  
 Db 2433 CATGGTGGGTAGGAAGATGTATATGATATGAATCTGGCCCCCAATTCATCTCT 2492  
 QY 540 ProGlyHisGlyHisArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGlnArg 559  
 Db 2493 CCAAGTCAAGGACATAGACTTCATCGACAGTACTCTCGGTCTTTGGTTTCAACGACAGA 2552  
 QY 560 HisProLeuGlnTyrSerProProIleHisGlyGlnGlnLeuProTyrGlyIleGln 579  
 Db 2553 CATCCACTCAGTACTCTCTCTCATTTATGACCAACACAGTTACCATATGTTATACAA 2612  
 QY 580 ArgValTyrArgHisSerProSerGluGluArgTyrLeuGlyLeuSerAsnGlnArgSer 599  
 Db 2613 AGGTTTACAGACATTCACCATCTGAAGAAAGATATTTGGGTTTATCCAATCAAAGGTCT 2672  
 QY 600 ProArgSerAsnSerSerLeuAspProLys 609  
 Db 2673 CCTCGCAGTAACATCATTAGACCCCAAA 2702

RESULT 6  
 AV198402 3261 bp DNA linear PLN 12-JUN-2003  
 AV198402 Arabidopsis thaliana ecotype KZ-9 FRIGIDA (FRI) gene, complete cds.  
 DEFINITION Arabidopsis thaliana ecotype KZ-9 FRIGIDA (FRI) gene, complete cds.  
 ACCESSION AV198402  
 VERSION AV198402.1 GI:31558912  
 KEYWORDS  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
 REFERENCE 1 (bases 1 to 3261)  
 AUTHORS Gazzani, S., Gendall, A.R., Lister, C. and Dean, C.  
 TITLE Analysis of the molecular basis of flowering time variation in  
 Arabidopsis accessions  
 JOURNAL plant Physiol. 132 (2), 1107-1114 (2003)  
 REFERENCE 2 (bases 1 to 3261)  
 AUTHORS Gazzani, S. and Dean, C.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-DEC-2002) Cell and Developmental Biology, John Innes  
 Center, Colney Lane, Norwich NR4 7UH, UK  
 FEATURES  
 Location/Qualifiers  
 1..3261  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:3702"  
 /chromosome="4"  
 /map="between mi51 and mi122"  
 /note="ecotype: KZ-9"  
 <505..>2816  
 /gene="FRI"  
 join(<505..1459,1853..2025,2115..>2816)  
 /gene="FRI"

```

Db      2373 GAAGCGGACTTCCAGAGAGATGTATACACACAGATTAACAGTCCAGGTTGTACCC 2432
QY      460 MetGluMetProProValThrSerSerTyrSerProIleTyrArgAspArgSerPhe 479
Db      2433 ATGGAATGCCACAGTAACTCTTCACTGATTTCTCTATCTACCGTGTAGAGGTTT 2492
QY      480 ProSerGlnArgAspAspGlnAspGluLeuSerAlaLeuValSerSerTyrLeuGly 499
Db      2493 CTTAGTCAAGAGAGATGATGACCAAGATGAATATACAGTCTTGTGAGTAGTACTCGGC 2552
QY      500 ProSerThrSerPheProHisArgSerArgArgSerProGluTyrMetValProLeuPro 519
Db      2553 CCGTCAACATCTTTCTCATCGCTCAAGAAGATCCCGGAATATATGGTTCCACTTCCA 2612
QY      520 HisGlyGlyLeuGlyArgSerValTyrAlaTyrGluHisLeuAlaProAsnSerTyrSer 539
Db      2613 CATGGTGGGTTAGGAAGATGTATGATCATATGACATCTGGCCCAATTCATATCT 2672
QY      540 ProGlyHisGlyHisArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGlnArg 559
Db      2673 CCAGGTCACGGACATAGACTTTCATCGACAGTACTCTCGTCTTTGGTTACGGACAGAGA 2732
QY      560 HisProLeuGlnTyrSerProIleHisGlyGlnGlnGlnLeuProTyrGlyIleGln 579
Db      2733 CATCCACTACACTCTCTCTCCATTCATGACCAACACAGTTACCATATGGTATACAA 2792
QY      580 ArgValTyrArgHisSerProSerGluGluArgTyrLeuGlyLeuSerAsnGlnArgSer 599
Db      2793 AGGTTTACAGACATTCACCACTCTGAGAGAGATATTTGGTTTATCCAAATCAAGTCT 2852
QY      600 ProArgSerAsnSerSerLeuAspProIys 609
Db      2853 CCTCGCAGTAACATCATCATTAGACCCCAA 2882

RESULT 5
AY198403
LOCUS   Arabidopsis thaliana ecotype Kondara FRIGIDA (FRI) gene, complete
DEFINITION
cds.
ACCESSION AY198403
VERSION   AY198403.1 GI:31558914
KEYWORDS
SOURCE   Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
REFERENCE 1 (bases 1 to 3150)
AUTHORS   Gazzani, S., Gendall, A.R., Lister, C. and Dean, C.
TITLE     Analysis of the molecular basis of flowering time variation in Arabidopsis accessions
JOURNAL   Plant Physiol. 132 (2), 1107-1114 (2003)
REFERENCE 2 (bases 1 to 3150)
AUTHORS   Gazzani, S. and Dean, C.
TITLE     Direct Submission
JOURNAL   Submitted (17-DEC-2002) Cell and Developmental Biology, John Innes Center, Colney Lane, Norwich NR4 7UH, UK
FEATURES
     source
     1..3150
         /organism="Arabidopsis thaliana"
         /mol_type="genomic DNA"
         /db_xref="taxon:3702"
         /chromosome="4"
         /map="between ml51 and ml122"
         /note="ecotype: Kondara"
         <394..>2705
         /gene="FRI"
         join(<394..1348,1742..1914,2004..>2705)
         /gene="FRI"
         /product="FRIGIDA"
         join(394..1348,1742..1914,2004..2705)
         /gene="FRI"

```

```

/note="flowering gene"
/codon_start=1
/product="FRIGIDA"
/protein_id="AAP49809.1"
/db_xref="GI:31558915"
translation="MSNYPTVAQAQPTTIANPLLQHQSEQRRELPKIVETESTSMD
ITICQKQPOLIKSIDELAAFAVETFRKQFDLQKHIESIENADLSKLENGVIA
ANNFNQPMLSPPNNVSVETTVTSQPSQEIPTTSNKPEGRWCELMCSKGLRKY
IYANISQAKLMBEIPSAKLAKPEVLDICIKFYLGRRFATKESMSARQVSL
LILSSFLMDRGGKVKIESWKDBAETAAVAWRKRLNTEGGIAAAEKMDARGLILL
VACGVPSPRSTDLDDLIMSGSNETAGALKESQFLVPMVSGIVESSIKRGVHIEAL
EMVTFQWEXKPSAALVITSLXSKSESFERAKKQASPLAKPEATKOLAVLSVWQ
CMETHFDPAKELPGWQIKQIVLSERKDLQDKEMEERKASLSLMEERAAKRMVQ
IKKPRISPMEMPEVTSVPIYRDRSPFSQDDQDELSALVSYLGPSTSPFPHRS
RRSEYVNPVPHGGGLGRSVYAEHLAPNSYSPGHLRHQVSPSLVHVGQRHLPQVSP
PIHQOQLPYGIQVYRHSPEERYVLGLSNQSPRSPNSLDPK"
BASE COUNT 951 a 516 c 645 g 938 t
ORIGIN
Alignment Scores:
Pred. No.: 1,27e-186 Length: 3150
Score: 2975.00 Matches: 607
Percent Similarity: 78.83% Conservative: 0
Best Local Similarity: 78.83% Mismatches: 2
Query Match: 95.84% Indels: 162
DB: 8 Gaps: 2
US-09-890-475-1 (1-609) x AY198403 (1-3150)
QY 1 MetSerAsnTyrProThrValAlaAlaGlnProThrThrAlaAsnProLeuLeu 20
Db 394 ATGTCCATTAATCCACGACGCGTGGCGGCAACCCACACGACGGCGAATCCACTGCTG 453
QY 21 GlnArgHisGlnSerGluGlnArgArgGluLeuProIysIleValGluThrGluSer 40
Db 454 CAGCGACATCAATCTGAACAGCAGCAGAGAAATACCGAAGATTGTGGAACAGAGTCT 513
QY 41 ThrSerMetAspIleThrIleGlyGlnSerIysGlnProGlnPheLeuLysSerIleAsp 60
Db 514 ACAAGTATGGACATTCAGATCGGTCAATCTAAGCAGGCTCAATTTTGAATCCATAGAC 573
QY 61 GluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAspLeuGln 80
Db 574 GAATTAGCTCGCTTTTCAGTTGCGAGTGAAACATTCAAACGCCAATTCGATGATCTTCAG 633
QY 81 LysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsnGlyValVal 100
Db 634 AAGCAGATCGAGTCAATCGAAACGCAATTAATTCCAAACCTCGAGAGTAACGGGTTGTC 693
QY 101 LeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSerProArgAsnAsnVal 120
Db 694 CTCGGCGCGCGAACAATAATTTCCATCAGCCGATGTTATCGCCTCGCGGAACAATGTA 753
QY 121 SerValGluThrThrValThrValSerGlnProSerGlnGluIleValProGluThrSer 140
Db 754 TCTGTAGAAACACCGCTCAGTGTAGCCACCGCTCTCAGGAGATTGTACCGAGACGCTG 813
QY 141 AsnLysProGluGlyGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyr 160
Db 814 AATAACCGGAGGGGAGCGTATGTGTGAGTTGATGTGTACAAAGTCTGCGTAATATAC 873
QY 161 IleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGluGluIleProSerAlaLeuLys 180
Db 874 ATATACGCAATATCTCTGATCAAGCTAAGTTAATGGAAGAGATTCCTTCCAGCTTTGAAA 933
QY 181 LeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPheTyrLeuGlnGly 200
Db 934 TTGCCCAAGAGACCGAGAGTTTGTATGTATGTATGTGCAAGTTTACTTACAGGG 993
QY 201 ArgArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeulle 220
Db 994 CGTAGACATTTACTAAGAGTCGCTATGAGCTCTGCGAGACAAAGTTTCGCTTCTTATA 1053

```

```

/db_xref="GI:10801174"
/translation="WSNYPPVAAQPTTANPELLOPHCSORRELPKIVENESTSMD
ITIGSKQDQFLKSIDELAAARSVAVEFKQFDLQKHIESINADISKLESGVLA
ARNNFHQPMLSPFNNSVETITVVSQPSGEIVFSTNPEGRMCMLCKSLGRKY
IYANTSDQAKMEIEIPSKIAKAEKPFVLDICIGFKYQORRAFTKESPMSSARQVSL
LILESFLMPDRGKVKIEMWIDEAETAAVRKELMTEGGAAAEKMDARGLLL
VACFVPSNFSITDILDIRMSGNEIAGALKRSQFLVPMVSGIVVSSIKRGMHIEAL
EMVYTFGMEDKFSALVLVTSFKMSKESPERAKRKAQSPFAKBAATKOLAVLSVWQ
CWEHLDPAKELPMQWIKEOIVSELDKTDQDKMEEKARSLSLMEEAALAKRMVNO
CIKPELSPMBMPVWTSSVSPIDRSFPQRDDODEISALVSVLQPSNFSFPHRS
RASPYWVPLPHGGLGRSVAYEHLAPNSYSPGHGHLRQYSPFSLVHGQRHLQYSP
PIHQQQQPIGYQIRYRHSPEERYLGNQSRPSNSLDPK"
BASE COUNT      1174 a      717 c      775 g      1095 t
ORIGIN

```

```

Alignment Scores:
Pred. No.:      6 43e-187      Length:      3761
Score:          2981.00      Matches:      608
Percent Similarity: 78.96%      Conservative: 0
Best Local Similarity: 78.96%      Mismatches: 1
Query Match:      96.04%      Indels:      162
DB:              8      Gaps:      2

```

US-09-890-475-1 (1-609) x AF228499 (1-3761)

```

QY      1 MetSerAsnTyrProProThrValAlaAlaGlnProThrThrThrAlaAsnProLeuLeu 20
DB      574 ATGTCCAAATATCCACCGAGCGTGGCGCGCAACCCACACGACGCGGAATCCACTGCTG 633
QY      21 GlnArgHisGlnSerGluGlnArgArgGluLeuProLysIleValGluThrGluSer 40
DB      634 CAGCGACATCAATCTGAACAGCGACGAGAGAAATTACCGAGATTGTGGAACAGAGTCT 693
QY      41 ThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLysSerIleAsp 60
DB      694 ACAAGTATGCATACGATCGGTCAATCTAAGCAGCGCTCAATTTTGAAATCCATAGAC 753
QY      61 GluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAspAspLeuGln 80
DB      754 GAATTAGCTGCGGTTTTTCAGTTCAGTGGAAACATTTCAACGCCAATTCGATGATCTTCAG 813
QY      81 LysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsnGlyValVal 100
DB      814 AAGCACATCGATCAATCGAAACCAATTCATTCCAAACTCGAGAGTAACCGCGGTGTCTC 873
QY      101 LeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSerProArgAsnAsnVal 120
DB      874 CTCGCCGCGGGAACAATAATTTCCATCAGCGGATGTTATCGCTCCCGGGAACAATGTA 933
QY      121 SerValGluThrThrValThrValSerGlnProSerGlnGluIleValProGluThrSer 140
DB      934 TCTGTAGAAACCCACCGTCACCTGTGAGCCAAACCGTCTCAGGAGATTGTATCCGAGACGTCG 993
QY      141 AsnLysProGluGluGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyr 160
DB      994 AATAAACCGGAGGGGAGCGTATGTGTGAGTTGATGTGTAGCAAGGTCTCGCTAAATATC 1053
QY      161 IleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGluGluIleProSerAlaLeuLys 180
DB      1054 ATATACCGGAATATCTGTATCAAGCTAAGTTAATGGAGAGATTCCTTCAGCTTGAAA 1113
QY      181 LeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPheTyrLeuGlnGly 200
DB      1114 TTGGCCCAAGGAGCCAGCAAGTTTGTATTTGATTGTATTTGGCAAGTTTACTTCAAGGG 1173
QY      201 ArgArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuIle 220
DB      1174 CGTAGAGCATTTACTTAAGAGTCGCGCTATGAGTCTCGGAGACAAAGTTTCGCTTCTATA 1233
QY      221 LeuGluSerPheLeuLeuMetProAspArgGlyLysGlyLysValLysIleGluSerTrp 240
DB      1234 CTGGAGTCTTTCTTCTAATGCTGATCGTGTAAAGGGAAGGTCAAGATTGAGAGTTGG 1293

```

```

QY      241 IleLysAspGluAlaGluThrAlaAlaValAlaTrpArgLysArgLeuMetThrGluGly 260
DB      1294 ATTTAAAGATGAGCGGAGACGCGTCTGTGTGTTGGAGGAAAGGTTGATGACTGAAGA 1353
QY      261 GlyLeuAlaAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuValAlaCysPhe 280
DB      1354 CGATTAGCTCGCGCTGAGAAATCGATGCAAGGGGTTTCTTTTACTAGTCTGCTGTTT 1413
QY      281 GlyValProSerAsnPheArgSerThrAspLeuLeuAspLeuIleArgMetSerGlySer 300
DB      1414 CGTGTCTCTCAACTTTAGAGTACAGATTGCTGCAATTGATAAGGATGAGTGTCTG 1473
QY      301 AsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetValSer----- 318
DB      1474 AATGAGATTGCGGTGCTTTTCAAGCGGTACAGTTTCTTTGTCCTATGCTCTCAGGTACC 1533
QY      318 ----- 318
DB      1534 ATATTCTGTTCTACTCGGTGAATTTCAITGCAAAAGTGGTCTTCTTTGTGACATCATC 1593
QY      318 ----- 318
DB      1594 GACCAACATCAAGTTCCATCTTCTGTTTTCGATAAGCTTGATGGTATAAACTAGGAGAGC 1653
QY      318 ----- 318
DB      1654 ACATCAAAATTTAGAGTGCATGACTGATTGAGCCAAATCCCTAGTAGAATTAATCTG 1713
QY      318 ----- 318
DB      1714 GAAAGAACTTGGAACTCTCAACCATAGGTTTGGTACGAAATTTGCTTGTTCAGAACCA 1773
QY      318 ----- 318
DB      1774 AATGATAGGCTATTGCGCTTGAAATAGTGTTTCTGTTTCCAAATATTGGAAGTTAAAA 1833
QY      318 ----- 318
DB      1834 TCGTATGATTAGCTGTTGGATACATAATTAAAGCTTAAGCAATCCCAACTCTAAGAAGTGG 1893
QY      319 ----- GlyIleValGluSerSerIleLysArgGlyMet 329
DB      1894 TACTTACACAAATATTCTATTGGTTCATAGGTATAGTTGAATCAAGTATCAAGCGTGAAGT 1953
QY      330 HisIleGluAlaLeuGluMetValTyrThrPheGlyMetGluAspLysPheSerAlaAla 349
DB      1954 CATATTGAGCTCTTGAGATGGTTTATACCTTTGGCATGGAGGATAAGTTTTCAGTGTCT 2013
QY      350 LeuValLeuThrSerPheLeuLysMetSerLysGluSerPheGluArgAlaLysArgLys 369
DB      2014 CTAGTTCTTAATCTTCACTTTAAAGATGAGCAAGAGTCAATTTGAGAGGGCAAAACGAAA 2073
QY      370 AlaGlnSerProLeuAlaPheLys----- 377
DB      2074 GCCCAGTCCCGCTGGCATTTGT-ATGAACCTTCCCTTGCACATTATGTACCTTTATGA 2132
QY      378 ----- GluAla 379
DB      2133 ACTCTTTATCATCTCTGAGTCTGACCATTTGATATATTTTCTCAACAGAAAGAGCG 2192
QY      380 AlaThrLysGlnLeuAlaValLeuSerSerValMetGlnCysMetGluThrHisLysLeu 399
DB      2193 GCTACAAAGCAGCTAGCTGTGTTATCATCATGTTATGTCAGTGTATGAGAGCTCAAGTTA 2252
QY      400 AspProAlaLysGluLeuProGlyTrpGlnIleLysGluGlnIleValSerLeuGluLys 419
DB      2253 GATCTCTGAAAGAACTACCAGGATGCCAGATCAAGAGCAAAATTTGTTAGTTGGAGAAA 2312
QY      420 AspThrLeuGlnLeuAspLysGluMetGluLysAlaArgSerLeuSerLeuMetGlu 439
DB      2313 GACACTCTTCAGCTCGAAAGAGATGGAAGAGAAAGCAAGATCTCTCAGTTTAATGGAG 2372
QY      440 GluAlaAlaLeuAlaLysArgMetTyrAsnGlnGlnIleLysArgProArgLeuSerPro 459

```

1354 GGATTAGCTGGCTGAGAAATGATGATGACAGGGTTTGTCTTTTACTAGTTGCTTGTGTTT 1413  
Db  
281 GlyValProSerAsnPheArgSerThrAspLeuLeuAspLeuLeuIleArgMetSerGlySer 300  
Qy  
1414 GGTGTTCCTTCAAACTTTAGAGTACAGATTGTCTGGATTGTGATTAAGATGAGTGTTCG 1473  
Db  
301 AsnGluIleAlaGlyAlaLeuIleArgSerGlnPheLeuValProMetValSer----- 318  
Qy  
1474 AATGAGATTGCGGTGCTTTGAGGGGTACAGTTTCTTGTCCCTATGGTCTCAGGTACC 1533  
Db  
318 ----- 318  
Qy  
1534 ATATTCTGTCTCTACTCGGTGAATTCATTCAAAAGGTGTTCCCTTTTGTGATCATC 1593  
Db  
318 ----- 318  
Qy  
1594 GACCAACATCAAGTCCATCTTTTGTTCGATAGCTTGATGATATAAAGTAGGAGC 1653  
Db  
318 ----- 318  
Qy  
1654 ACATCAATATTAGAGTCAATGACTGATTCAGCCAAATCCTAGCTAGAAATTAATCTG 1713  
Db  
318 ----- 318  
Qy  
1714 GAAAGAACTTGGAACCTCAACCATAGTTTTTGTACGAAATTTGCTTGTTCAGAACCA 1773  
Db  
318 ----- 318  
Qy  
1774 AATGATAGCTATTGCTTGAATAGTGTCTTGTGTTTCCCAATATTGGAAGTTAAAA 1833  
Db  
318 ----- 318  
Qy  
1834 TCGTATGACTTAGCTGTGGATACATAAAGCTTAAGCAATGCAACTCTAAGAAAGTGG 1893  
Db  
319 -----GlyIleValGluSerSerIleLysArgGlyMet 323  
Qy  
1894 TACTTACCAATATTCTATTGTCATAGTATAGTTGAATCAAGTATCAAGCGTGGATG 1953  
Db  
330 HisIleGluAlaLeuMetValThrPheGlyMetGluAspLysPheSerAlaAla 349  
Qy  
1954 CATATTGAAGCTCTTGAGATGGTTTATACCTTGGCATGGAGGATAAGTTTTCAGCTGCT 2013  
Db  
350 LeuValLeuThrSerPheLeuLysMetSerLysGluSerPheGluAlaLysArgLys 369  
Qy  
2014 CTAGTTCTTACCTTCATTCTTAAGATGACGAGAGTCACTTTGAGGGCAAAACGAAA 2073  
Db  
370 AlaGlnSerProLeuAlaPheLys----- 377  
Qy  
2074 GCCAGTCAACGCTGGCAATTTGT-ATGAACCCCTCCCTTGCACTATTATGATGA 2132  
Db  
378 -----GluAla 379  
Qy  
2133 ACTCTTTATCATCTGAGTCTGACCATGATGATATATTTATTTCTCAACAGAAAGACG 2192  
Db  
380 AlaThrLysGlnLeuAlaValLeuSerSerValMetGlnCysMetGluThrHisLysLeu 399  
Qy  
2193 GCTTACAAAGCAGTAGCTGTGTTATCATCAGTTATGCTAGTGTATGGAGACTCACAGTTA 2252  
Db  
400 AspProAlaLysGluLeuProGlyTrpGlnIleLysGluGlnIleValSerLeuGlyLys 419  
Qy  
2253 GATCCTGGGAGAACTACACAGATGGCAGATCAAGAGCAAAATGTTAGCTTGGAGAAA 2312  
Db  
420 AspThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSerLeuSerLeuMetGlu 439  
Qy  
2313 GACACTCTTCAGCTCGACAAAGAGATGGAAGAGAAAGCAAGATCTCTCAGTTTAATGGAG 2372  
Db  
440 GluAlaAlaLeuAlaLysArgMetThrAsnGlnGlnIleLysArgProArgLeuSerPro 459  
Qy  
2373 GAAGCCGCACTTGCACAGAGATGTATACCAACAGATATAAACCGTCCAGGTTCGACCC 2432  
Db  
460 MetGluMetProProValThrSerSerSerThrProIleLeuThrArgAspArgSerPhe 479  
Qy  
2433 ATGGAATAATGCCACCAGTAATCTTCTTCATCGTATTCTCTATCTACCGTGATAGAGCTTT 2492  
Db

480 ProSerGlnArgAspAspGlnAspGluIleSerAlaLeuValSerSerThrLeuGly 499  
Qy  
2493 CTTAGTCAAGAGACGATGACCAAGATGAATATCAGCTCTTGTGATGATGTTACCTCGGC 2552  
Db  
500 ProSerThrSerPheProHisArgSerArgArgSerProGluThrMetValProLeuPro 519  
Qy  
2553 CCGTCAACATCTTTTCTCATCTCCTCAAGAAGATCCCGGAATATATGTTCCACTTCCA 2612  
Db  
520 HisGlyGlyLeuGlyArgSerValTyAlaTyGluHisLeuAlaProAsnSerThrSer 539  
Qy  
2613 CATGTTGGTTAGGAAGATGTATATGATATGAACATCTGGCCCCCAATCATACTCT 2672  
Db  
540 ProGlyHisGlyHisArgLeuHisArgGlnTySerProSerLeuValHisGlyGlnArg 559  
Qy  
2673 CCAGTCCAGGACATAGACTTCATCGACAGTACTCTCCGCTTTGGTTCACGACAGAGA 2732  
Db  
560 HisProLeuGlnTySerProProIleHisGlyGlnGlnGlnLeuProTyGlyIleGln 579  
Qy  
2733 CATCCACTACAGTACTCTCTCCAAATTCATGGACACACAGTTACATATGTTATACAA 2792  
Db  
580 ArgValTyArgHisSerProSerGluGluArgTyLeuGlyLeuSerAsnGlnArgSer 599  
Qy  
2793 AGGTTTACACATTCACCATCTGAAGAAGATATTGGGTTTATCCAAATCAAGGTCT 2852  
Db  
600 ProArgSerAsnSerSerLeuAspProLys 609  
Qy  
2853 CCTCGCAGTAATCATCTAGACCCCAA 2882  
Db  
RESULT 4  
AF228499 3761 bp DNA linear PLN 14-OCT-2000  
LOCUS Arabidopsis thaliana FRIGIDA (FRI) gene, complete cds.  
DEFINITION Arabidopsis thaliana FRIGIDA (FRI) gene, complete cds.  
ACCESSION AF228499  
VERSION AF228499.1 GI:10801173  
KEYWORDS  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
REFERENCE 1 (bases 1 to 3761)  
AUTHORS Johanson,U., West,J., Lister,C., Michaels,S., Amasino,R. and Dean,C.  
TITLE Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time  
JOURNAL Science 290 (5490), 344-347 (2000)  
MEDLINE 20485641  
PubMed 11030654  
REFERENCE 2 (bases 1 to 3761)  
AUTHORS Johanson,U. and Lister,C.  
TITLE Direct Submission  
JOURNAL Submitted (25-JAN-2000) Molecular Genetics, John Innes Centre, Colney Lane, Norwich, Norfolk NR4 7UH, U.K.  
FEATURES  
source  
1..3761  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/cultivar="ecotype H51"  
/db\_xref="taxon:3702"  
/chromosome="4"  
/map="between msi and m122"  
gene  
574..>2885  
/gene="FRI"  
mRNA  
/gene="FRI"  
/product="FRIGIDA"  
CDS  
join(574..1528,1922..2094,2184..>2885)  
/gene="FRI"  
/note="late flowering gene"  
/codon\_start=1  
/product="FRIGIDA"  
/protein\_id="AAG23414.1"

341 GlyMetGluAspLysPheSerAlaAlaLeuValLeuThrSerPheLeuLysMetSerLys 360  
 Db GGATGGAGGATAGTTTTCAGCTGCTAGTTCCTTCACTTCACTTAAAGATGAGCAAG 1441  
 361 GluSerPheGluArgAlaLysArgLysAlaGlnSerProLeuAlaPheLysGluAlaAla 380  
 Db GAGTCATTTGAGAGGGCAAAACGAAAGCCCGAGTCCCGCTGGCATTTTAAAGAGCGGCT 1501  
 381 ThrLysGlnLeuAlaValLeuSerSerValMetGlnCysMetGluThrHisLysLeuAsp 400  
 Db ACAAAGCAGTACTGTTTATCATCAGTTATGCGAGTGTATGAGACTCACAGTTAGAT 1561  
 401 ProAlaLysGluLeuProGlyTrpGlnLysGluGlnLeuValSerLeuGluLysAsp 420  
 Db CTTGCGAAGAACTACAGGATGCGAGATCAAGAGCAAAATTTAGTGTGAGAGAAC 1621  
 421 ThrLeuGlnLeuAspLysGluMetGluGluLysAlaArgSerLeuSerLeuMetGluGlu 440  
 Db ACTCTTCAGCTCGACAAAGAGATGGAGAGAAAGCAAGATCTCTCAGTTTAAATGGAGAA 1681  
 441 AlaAlaLeuAlaLysArgMetTyrAsnGlnGlnLysLeuArgProArgLeuSerProMet 460  
 Db GCGCACTTCCCAAGAGATGTTATACCAACAGATAAAACGTCCTCAAGGTTGTCAACCATG 1741  
 461 GluMetProValThrSerSerTyrSerProLysTyrArgAspArgSerPhePro 480  
 Db GAAATGCCACGTAATCTTCTCATCGTATTTCTCTATCTACCGTATAGAAAGCTTCTCT 1801  
 481 SerGlnArgAspAspGlnAspGluLysSerAlaLeuValSerSerTyrLeuGlyPro 500  
 Db AGTCAAAGACGATGACCAAGATGAATATCAGCTCTTGTGAGTAGTTTACCTCGGCCCG 1861  
 501 SerThrSerPheProHisArgSerArgSerProGluTyrMetValProLeuProHis 520  
 Db TCAACATCTTTTCTCATCTCCATCAAGAGATCCCGGAATATATGTTTCCATCTCCACAT 1921  
 521 GlyGlyLeuGlyArgSerValTyrAlaTyrGluHisLeuAlaProAsnSerTyrSerPro 540  
 Db GTGGGTTAGGAAGATGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1981  
 541 GlyHisGlyHisArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGlnArgHis 560  
 Db GGTCAAGGACATAGACTTCTACAGAGTACTCTCCGCTTTGGTTCCAGGACAGAGACAT 2041  
 561 ProLeuGlnTyrSerProLysHisGlyGlnGlnLeuProTyrGlyTyrGlnArg 580  
 Db CCATCAGTACTCTCTCCATTCATTCAGCAACACAGATTACCATATGTTATACAAAGG 2101  
 581 ValTyrArgHisSerProSerGluGluArgTyrLeuGlyLeuSerAsnGlnArgSerPro 600  
 Db GTTTACAGACATTCCCATCTGAGAAAGATATTGGTTTATCCCAATCAAAGGTCTCTCT 2161  
 601 ArgSerAsnSerSerLeuAspProLys 609  
 Db CGCAGTAACCTCATCTAGACCCCAAA 2188  
 RESULT 3  
 LOCUS AX032762 3761 bp DNA linear PAT 21-SEP-2000  
 DEFINITION Sequence 2 from Patent WO0046358.  
 ACCESSION AX032762  
 VERSION AX032762.1 GI:10279738  
 KEYWORDS Arabidopsis thaliana (thale cress)  
 SOURCE Arabidopsis thaliana  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
 1  
 REFERENCE Dean, C., West, J. and Johanson, J.  
 AUTHORS Plant Gene  
 TITLE Patent: WO 0046358-A 2 10-AUG-2000;  
 JOURNAL DEAN CAROLINE (GB); WEST JOANNE (GB); PLANT BIOSCIENCE LTD (GB);

JOHANSON URBAN (SE)  
 Location/Qualifiers  
 1. 3761  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:3702"  
 BASE COUNT 1174 a 717 c 775 g 1095 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 6,43e-187 Length: 3761  
 Score: 2981.00 Matches: 608  
 Percent Similarity: 78.96% Conservative: 0  
 Best Local Similarity: 78.96% Mismatches: 1  
 Query Match: 96.04% Indels: 162  
 DB: Gaps: 2  
 US-09-890-475-1 (1-609) x AX032762 (1-3761)  
 QY 1 MetSerAsnTyrProThrValAlaAlaGlnProThrThrAlaAsnProLeuLeu 20  
 Db 574 ATGTCCAAATTAATCCACCGAGCGTGGCGGCAACCCACAGCGGCGAATCCACTCTG 633  
 QY 21 GlnArgHisGlnSerGluGlnArgArgGluLeuProLysIleValGluThrGluSer 40  
 Db 634 CAGCGACATCAATCTGAACAGCGACGAAGAGAAATTCGGAAGATTGTCGAACACAGATCT 693  
 QY 41 ThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLysSerIleAsp 60  
 Db 694 ACAAGTATCGCATTCAGCATCGTCAATCTAAGCAGCCCTCAATTTTGAATCCATAGAC 753  
 QY 61 GluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAspAspLeuGln 80  
 Db 754 GAATTAGCTGGCTTTTTCAGTTGAGTGGAAACATTCAAACGCCAATTCGATGATCTTCAG 813  
 QY 81 LysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsnGlyValVal 100  
 Db 814 AAGCACATCGAGTCAATCGAAACCGCAATTCATCCAAACCTCGAGAGTAACGGGTGTC 873  
 QY 101 LeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSerProProArgAsnAsnVal 120  
 Db 874 CTGCGCGCGGACACATATATTTCCATCAGCGGATGTTATCGCTCCGCGGAACAATGTA 933  
 QY 121 SerValGluThrThrValThrValSerGlnProSerGlnGlnIleValProGluThrSer 140  
 Db 934 TCTGTAGAAACCACTCTCTGAGCCACCGCTCTCAGGAGATTGTACCGGAGACGTCG 993  
 QY 141 AsnLysProGluGlyGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyr 160  
 Db 994 AATAAACCGGAGGGGAGCGTATGTGTAGTTGATGTGTAGCAAGGTCTCGTAATATAC 1053  
 QY 161 IleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGluGluIleProSerAlaLeuLys 180  
 Db 1054 ATATACGCAATATCTCTGATCAAGCTAAGTTAATGGAAGAGATTCTTTCAGCTTGAAA 1113  
 QY 181 LeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPheTyrLeuGlnGly 200  
 Db 1114 TTGGCAAGGAGCGACGCAAGTTTGTATTGGATTGTATTGGCAAGTTTACTTACAGGG 1173  
 QY 201 ArgArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeu 220  
 Db 1174 CGTAGACATTTACTTAAGAGTCGCTATGAGCTCTCGAGACAAAGTTTCGCTTCTTATA 1233  
 QY 221 LeuGluSerPheLeuLeuMetProAspArgGlyLysGlyLysValIleLeuSerTrp 240  
 Db 1234 CTGGAGTCTTTTCTTAAATGCTGTATCGTGTAAAGGAAAGGTGAAGATTGAGATTGG 1293  
 QY 241 IleLysAspGluAlaGluThrAlaAlaValAlaTrpArgLysArgLeuMetThrGluGly 260  
 Db 1294 ATTAAAGATGAGCGGAGACGGCTGCTGTGCTTGGAGGAAAGGTTGATGACTCAAGGA 1353  
 QY 261 GlyLeuAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuValAlaCysPhe 280

Db	1201	CCTGCGAAGAACTACCAAGGATGGCAGATCAAGAGCAAAATGTAGCTGGAGAAGAC	1260
Qy	421	ThrLeuGlnLeuAspLysGluMetGluGlyLysAlaArgSerLeuSerLeuMetGluGlu	440
Db	1261	ACTCTTCAGCTCGACAAGAGATGGAAGAGAAAGCAAGATCTCTCAGTTTAATGAGGAA	1320
Qy	441	AlaAlaLeuAlaLysArgMetTyrAsnGlnGlnLysArgProArgLeuSerProMet	460
Db	1321	GCCGCACCTTCCCAAGAGAATGTATACCAACAGATAAACCTCCAAAGTTGTCCACCATG	1380
Qy	461	GluMetProProValThrSerSerTyrSerProIleTyrArgAspArgSerPhePro	480
Db	1381	GAATGCCACAGTAACTCTTCATCGTATCTCTCTATCTACCGTGATAGAGTTTCTCT	1440
Qy	481	SerGlnArgAspAspGlnAspGluLysSerAlaLeuValSerSerTyrLeuGlyPro	500
Db	1441	AGTCAAGAGAGCGATGACCAAGATGAAATATCAGCTCTTGTGAGTAGTTACTCGGCCG	1500
Qy	501	SerThrSerPheProHisArgSerArgSerProGluTyrMetValProLeuProHis	520
Db	1501	TCAACATCTTTTCTCATCTCGCTCAAGAGATCCCGGATATATGGTTCCACITCCACAT	1560
Qy	521	GlyGlyLeuGlyArgSerValTyrAlaTyrGluHisLeuAlaProAsnSerTyrSerPro	540
Db	1561	GGTGGGTTAGGAAGAAGTGTATATCATATGAACATCTGGCCCAAAATTCATCTCTCCA	1620
Qy	541	GlyHisGlyHisArgLeuHisArgGlnTyrSerProSerLeuValHisGlyGlnArgHis	560
Db	1621	GGTCAGGACATAGACTTCATCAGCAGTACTCTCCGTCTTTGGTTCCAGGACAGACAT	1680
Qy	561	ProLeuGlnTyrSerProProIleHisGlyGlnGlnLeuProTyrGlyLeuArg	580
Db	1681	CCACTACGACTCTCTCTCCAAATTCATGACAAACACAGTTACCATATGGTATACAAAGG	1740
Qy	581	ValTyrArgHisSerProSerGluGlnArgTyrLeuGlyLeuSerAsnGlnArgSerPro	600
Db	1741	GTATTACAGACATTCACCATCTGAAGAAAGATATTTGGGTTTATCCAAATCAAGGTCCT	1800
Qy	601	ArgSerAsnSerSerLeuAspProLys	609
Db	1801	CGCAGTAACTCATATAGACCCCAAA	1827

RESULT 2  
AX032763  
LOCUS 2257 bp DNA linear PAT 21-SEP-2000  
DEFINITION Sequence 3 from Patent WO0046358.  
ACCESSION AX032763  
VERSION AX032763.1 GI:10279739  
KEYWORDS synthetic construct  
SOURCE synthetic construct  
ORGANISM artificial sequences.  
REFERENCE 1  
AUTHORS Dean C., West, J. and Johanson, U.  
TITLE Plant gene  
JOURNAL Patent: WO 0046358-A 3 10-AUG-2000;  
DEAN CAROLINE (GB); WEST JOHANN (GB); PLANT BIOSCIENCE LTD (GB);  
JOHANSON URBAN (SE)  
FEATURES  
source  
1..2257  
/organism="synthetic construct"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32630"  
/note="likely cDNA sequence of the H51 FRI gene"  
BASE COUNT 698 a 443 c 493 g 623 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 2 58e-195 Length: 2257  
Score: 3104.00 Matches: 609  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0

Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0
US-09-890-475-1 (1-609) x AX032763 (1-2257)			
Qy	1	MetSerAsnTyrProProThrValAlaAlaGlnProThrThrAlaAsnProLeuLeu	20
Db	362	ATGTCATATTCACCGAGCGGTGGCGCGCAACCCACACGCGCGAATCCACTGCTG	421
Qy	21	GlnArgHisGlnSerGluGlnArgArgGluLeuProLysIleValGluThrGluSer	40
Db	422	CAGCGACATCAATCTGAACACGCGACGAGAGAAATACCGAAGATTGTCCGAACAGAGTCT	481
Qy	41	ThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLysSerIleAsp	60
Db	482	ACAAGTAGGACATACGATCGGTCAATCAAGCAGCGCTCAATTTTGAATCCATAGAC	541
Qy	61	GluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAspLeuGln	80
Db	542	GAATTAGCTGCTTTTCAGTTTCAGTTCAGTGAACATCAACCGCAATTCGATGATCTTCAG	601
Qy	81	LysHisIleGluSerIleGluAsnAlaIleAspSerLysLeuGluSerAsnGlyValVal	100
Db	602	AGCACATCGAGTCAATCGAAGACGCAATTTGATTCCAACTCGAGAGTAAACGGCTGTC	661
Qy	101	LeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSerProProArgAsnAsnVal	120
Db	662	CTCGCGCGCGGAGCAATAATTCATCAGCCGATGTATCGCTCCGCGGAAACAATGTA	721
Qy	121	SerValGluThrThrValThrValSerGlnProSerGlnGluLeuValProGluThrSer	140
Db	722	TCTGTAGAAACCCCGCTCAGTGTAGCCACCCGCTCAGGAGATTGTACCGGAGACGTCG	781
Qy	141	AsnLysProGluGlyGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyr	160
Db	782	ATAAACCAGGAGGGGAGCTATGTGTAGTTGATGTAGTACCAAGGCTCGCGTAATAC	841
Qy	161	IleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGluGluLeuProSerAlaLeuLys	180
Db	842	ATATACGCGAATATCTCTGATCAAGCTAAGTTAATGGAAGAGATTCCTTCAGCTTGAAA	901
Qy	181	LeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPheTyrLeuGlnGly	200
Db	902	TTGGCCACAGGAGCCAGCGAAGTTTGTATTGATTGTATTGGCAAGTTTACTTACAAGG	961
Qy	201	ArgArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuLe	220
Db	962	CGTAGAGCAATTTACTAAAGAGTCGCTATGAGCTCTCGAGACAAAGTTTCGCTTCTATA	1021
Qy	221	LeuGluSerPheLeuLeuMetProAspArgGlyLysGlyLysValLysIleGluSerTyr	240
Db	1022	CTGGAGTCTTTTCTTCTAATGCCGTGATCGTGTAAAGGAAAGGTTGAAGATTGAGATTG	1081
Qy	241	IleLysAspGluAlaGluThrAlaAlaValAlaTyrArgLysArgLeuMetThrGluGly	260
Db	1082	ATTAAAGATGAGCGGAGACGCGCTCTCTGCTGGAGAAAGGTTGATGACTGAAGGA	1141
Qy	261	GlyLeuAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuValAlaCysPhe	280
Db	1142	GGATTAGCTGGCGCTGAGAAATGGAATGGAAGGGTTTGTCTTTTACTAGTTGCTTTT	1201
Qy	281	GlyValProSerAsnPheArgSerThrAspLeuLeuAspLeuLeuArgMetSerGlySer	300
Db	1202	GGTGTCTCTTCAAACTTTAGGAGTACAGATTGCTGGATTTTGATAAGGATGAGTGGTTCG	1261
Qy	301	AsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetValSerGlyIle	320
Db	1262	AATGAGATTGCGCGTCTTTGAAGCGGTCACAGTTTCTTGTCCTTATGCTCTCAGTATA	1321
Qy	321	ValGluSerSerIleLysArgGlyMetHisIleGluAlaLeuGluMetValTyrThrPhe	340
Db	1322	GTGTGATCAAGTATCAAGCGTGAATGCATATTTGAAGCTCTTGAGATGGTTTATACCTTT	1381

```

AF228500      1830 bp      mRNA      linear      PLN 14-OCT-2000
LOCUS         Arabidopsis thaliana FRIGIDA (FRI) mRNA, complete cds.
DEFINITION
ACCESSION     AF228500
VERSION       AF228500.1  GI:10801175
KEYWORDS
SOURCE        Arabidopsis thaliana (thale cress)
ORGANISM      Arabidopsis thaliana
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE     1 (bases 1 to 1830)
AUTHORS       Johanson,U., West,J., Lister,C., Michaels,S., Amasino,R. and
               Dean,C.
TITLE         Molecular analysis of FRIGIDA, a major determinant of natural
               variation in Arabidopsis flowering time
JOURNAL       Science 290 (5490), 344-347 (2000)
MEDLINE       20485641
PubMed       11030654
REFERENCE     2 (bases 1 to 1830)
AUTHORS       Johanson,U. and Lister,C.
TITLE         Direct Submission
JOURNAL       Submitted (25-JAN-2000) Molecular Genetics, John Innes Centre,
               Colney Lane, Norwich, Norfolk NR4 7UH, U.K.
FEATURES
source        Location/Qualifiers
               1..1830
               /organism="Arabidopsis thaliana"
               /mol_type="mRNA"
               /cultivar="ecotype H51"
               /db_xref="taxon:3702"
               /chromosomes="4"
               /map="between mi51 and mi122"
gene          1..1830
               /gene="FRI"
CDS           1..1830
               /gene="FRI"
               /notes="late flowering gene"
               /codon_start=1
               /product="FRIGIDA"
               /protein_id="AAG23415.1"
               /db_xref="GI:10801175"
               /translation="MSNYEPTVAAGFTTANPLLRHQHSEQRRELPRKIVETESTMD
               ITGQSKQFLLKSIDELAAFSVAVETFKRFDDLRKHIESIENAIIDSKLNSGVLIA
               ARNNPHQMLSPNNVSVETVTSQPSOEIVPTSNKEPGRMCELMCKGLRY
               IVANTI SDOAKLMEETPSALKAKPAKFLVDCIGKFLQGRRAFTKSPMSARQVSL
               LILSEFLMPDRGKGVKIESWKDEASTRAVAKRLMTETGGLAAAEKXDAKGLILL
               VACGPVSNFRSTDLLIRMSGSENIAGALKRSQFLVPMVSGIVESIRGPHIEAL
               EMWTFGMDKFSALVLVTSFUMSKSEFERAKRAQSPFLAKKAATKQLAVLSSVMQ
               CMETHKLDPAKLPQWQKEQIVSLEKDTLQDKEMEERKARSLMEEAALAKEMYNQ
               QIKRPLSPMEPPVTSYSPIYDRSPFPQDDQDEISALVSYLGPSTPFRHS
               RSPETWVPLPHGGLGRSVYAYEHLAPNSYSPGHGRLHROYSPSLVHGQRHPLQYSP
               PTHGOOLPEYGIQVRVHSPSEERVLGLSNQSPRSNSSLDPK"
BASE COUNT   540 a 379 c 436 g 475 t
ORIGIN
Alignment Scores:
Pred. No.:    1..96e-195      Length:    1830
Score:         3104.00        Matches:    609
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels:    0
DB:             8            Gaps:      0

US-09-890-475-1 (1-609) x AF228500 (1-1830)

QY      1 MetSerAsnTyrProProThrValAlaAlaGlnProThrThrThrAlaAsnProLeuLeu 20
Db      1 ATGTCCAAATTATCCACCGAGGTGCGCGGCAACCCCAACCAACGACGCGCAATCCACTGCTG 60

QY      21 GlnArgHisGlnSerGluGlnArgArgGluLeuProLysIleValGluThrGluSer 40
Db      61 CAGCGACATCAATCTCTGAACAGCGAGCAAGAGAATTAACCGAAGATTGCGAAACAGAGTCT 120

```

```

41 ThrSerMetAspIleThrIleGlyGlnSerLysGlnProGlnPheLeuLysSerIleAsp 60
Db      121 ACARGTATGGACATTACGATCGTCAATTAAGCAGCGCTCAATTTTGAATTCCTATGAC 180

QY      61 GluLeuAlaAlaPheSerValAlaValGluThrPheLysArgGlnPheAspAspLeuGln 80
Db      181 GAATTAGCTGCGTTTTCAGTTGCAGTGGAACATTCAAACGCCCAATTTCGATGATCTTCAG 240

QY      81 LysHisIleGluSerIleGluAsnAlaIleAspSerLysLysLeuGluSerArgGlyValVal 100
Db      241 AAGCATCATCGATCAATCGAAACCAATGATTCCAAACTCGAGAGTAACGGGTGTC 300

QY      101 LeuAlaAlaArgAsnAsnAsnPheHisGlnProMetLeuSerProProArgAsnAsnVal 120
Db      301 CTCGCGCGCGCAACAATAATTTCCATCAGCGATGTTATCGCTCCGCGGAACAATGTA 360

QY      121 SerValGluThrThrValThrValSerGlnProSerGlnGluIleValProGluThrSer 140
Db      361 TCTGTAGAAACACCCGTCACGTGTGAGCCACCGTCTCAGGAGATTGTACCGGAGACGCTCG 420

QY      141 AsnLysProGluGlyGlyArgMetCysGluLeuMetCysSerLysGlyLeuArgLysTyr 160
Db      421 AATAAACCGAGGGGGGACGATGTGTGAGTTGATGTGTAGCAAGGCTCTGCGTAATATAC 480

QY      161 IleTyrAlaAsnIleSerAspGlnAlaLysLeuMetGluGluIleProSerAlaLeuLys 180
Db      481 ATATACGCGCAATATCTCTGATCAAGCTAAGTTAATGGAAGAGATTCCTCAGCTTTGAAA 540

QY      181 LeuAlaLysGluProAlaLysPheValLeuAspCysIleGlyLysPheTyrLeuGlnGly 200
Db      541 TTGGCCNAGGAGCCAGCAAGTTTGTATTGATTGATTGCGCAAGTTTACTTACAGGG 600

QY      201 ArgArgAlaPheThrLysGluSerProMetSerSerAlaArgGlnValSerLeuLeuIle 220
Db      601 CGTAGAGCATTTTACTTAAAGAGTCGCTATGAGCTCTCGAGACAAAGTTTCCTTCTTATA 660

QY      221 LeuGluSerPheLeuLeuMetProAspArgGlyLysGlyLysValLysIleGluSerTrp 240
Db      661 CTGGAGTCTTTTCTTAATGCCTGATCGTGTAACGGAAGGTGAAGATTGAGATTGG 720

QY      241 IleLysAspGluAlaGluThrAlaAlaValAlaTrpArgLysArgLeuMetThrGluGly 260
Db      721 ATAAAGATGAGCGGAGACGCGTCTGTTGCTTGGAGGAAAAGTTGATGACTGAAGGA 780

QY      261 GlyLeuAlaAlaGluLysMetAspAlaArgGlyLeuLeuLeuValAlaCysPhe 280
Db      781 GGATTAGCTGCGGCTGAGAAAATGGATGCAAGGGGTTTGCTTTTACTAGTTGCTGTGTTT 840

QY      281 GlyValProSerAsnPheArgSerThrAspLeuLeuAspLeuIleArgMetSerGlySer 300
Db      841 GGTGTTCTCTTCAAACTTTAGGAGTACAGATTGCTGGATTGTATAAGGATGAGTGGTTCG 900

QY      301 AsnGluIleAlaGlyAlaLeuLysArgSerGlnPheLeuValProMetValSerGlyIle 320
Db      901 AATGAGATTGCGGTCGCTTTTGAACGCGTCACAGTTTCTTCTCCCTATGGTCTCAGGTATA 960

QY      321 ValGluSerSerIleLysArgGlyMetHisIleGluAlaLeuGluMetValTyrThrPhe 340
Db      961 GTTGAATCAAGTATCAACGCTGGAAATGCATATGAAGCTCTTGAGATGGTTTATACCTTT 1020

QY      341 GlyMetGluAspLysPheSerAlaAlaLeuValLeuThrSerPheLeuLysMetSerLys 360
Db      1021 GGCATGAGAGATAAGTTTTCAGCTGCTCTAGTTCTTACTTCTTCTTAAAGATGAGCAAG 1080

QY      361 GluSerPheGluArgAlaLysArgLysAlaGlnSerProLeuAlaPheLysGluAlaAla 380
Db      1081 GAGTCATTTGAGAGGGCAAAACGGAAGCCAGTCACCGCTGGCANTTTAAGAAAGCGGCT 1140

QY      381 ThrLysGlnLeuAlaValLeuSerSerValMetGlnCysMetGluThrHisLysLeuAsp 400
Db      1141 ACAAGCAGCTAGCTGTGTTATCATCAGTTATGCAAGTGTATGAGACTCAAGATTAGAT 1200

QY      401 ProAlaLysGluLeuProGlyTrpGlnIleLysGluGlnIleValSerLeuGluLysAsp 420

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

DM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 25, 2003, 08:48:34 ; Search time 6375 Seconds  
(without alignments)  
3908.073 Million cell updates/sec

Title: US-09-890-475-1

Perfect score: 3104

Sequence: 1 MSNYPPTVAQPTTTANPL.....RYLGLSNQSPRSNSLDPK 609

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n model -DEV=xlh  
-Q=/cgn2/1/USPTO\_spool/US03890475/runat\_14082003\_090349\_19545/app\_query.fasta\_1.775  
-DB=genEmbl -QMT=rapap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTMT=ptc -NORM=ext -HEARSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09890475@cgn 1 1 3963 @runat\_14082003\_090349\_19545 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DPSLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl: 1: gb.ba: 2: gb.htg: 3: gb.in: 4: gb.om: 5: gb.ov: 6: gb.pat: 7: gb.ph: 8: gb.pl: 9: gb.pr: 10: gb.ro: 11: gb.sts: 12: gb.sy: 13: gb.un: 14: gb.vi: 15: em.ba: 16: em.fun: 17: em.hum: 18: em.in: 19: em.mu: 20: em.om: 21: em.ov: 22: em.pat: 23: em.ph: 24: em.pl: 25: em.ro: 26: em.sts: 27: em.un:

29: em.vi: 30: em.htg.hum: 31: em.htg.inv: 32: em.htg.other: 33: em.htg.mus: 34: em.htg.pln: 35: em.htg.rod: 36: em.htg.mam: 37: em.htg.vrt: 38: em.sy: 39: em.htgo.hum: 40: em.htgo.mus: 41: em.htgo.other:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3104	100.0	1830	8	AF228500 Arabidops
2	3104	100.0	2257	6	AX032763 Sequence
3	2881	96.0	3761	6	AX032762 Sequence
4	2881	96.0	3761	8	AF228499 Arabidops
5	2975	95.8	3150	8	AY198403 Arabidops
6	2975	95.8	3261	8	AY198402 Arabidops
7	2975	95.8	3739	8	AY198401 Arabidops
8	2969	95.7	3301	8	AY198405 Arabidops
9	2950	95.0	3758	8	AY198404 Arabidops
10	2936	94.6	91849	8	F6N23 Arabidops
11	2936	94.6	137975	8	ATCHR1V2 Arabidops
12	931	30.0	558	8	AY092691 Arabidops
13	931	30.0	558	8	AY092693 Arabidops
14	931	30.0	558	8	AY092691 Arabidops
15	931	30.0	558	8	AY092693 Arabidops
16	927	29.9	558	8	AY092677 Arabidops
17	927	29.9	558	8	AY092679 Arabidops
18	927	29.9	558	8	AY092694 Arabidops
19	927	29.9	558	8	AY092695 Arabidops
20	927	29.9	558	8	AY092696 Arabidops
21	921	29.7	558	8	AY092678 Arabidops
22	921	29.7	558	8	AY092680 Arabidops
23	921	29.7	558	8	AY092682 Arabidops
24	921	29.7	558	8	AY092684 Arabidops
25	921	29.7	558	8	AY092685 Arabidops
26	921	29.7	558	8	AY092686 Arabidops
27	921	29.7	558	8	AY092687 Arabidops
28	921	29.7	558	8	AY092688 Arabidops
29	921	29.7	558	8	AY092689 Arabidops
30	921	29.7	558	8	AY092690 Arabidops
31	921	29.7	558	8	AY092692 Arabidops
32	845	27.2	511	8	AY176673 Brassica
33	658	21.2	503	8	AY092437 Arabidops
34	658	21.2	503	8	AY092438 Arabidops
35	658	21.2	503	8	AY092439 Arabidops
36	658	21.2	503	8	AY092440 Arabidops
37	658	21.2	503	8	AY092441 Arabidops
38	658	21.2	503	8	AY092442 Arabidops
39	658	21.2	503	8	AY092443 Arabidops
40	658	21.2	503	8	AY092444 Arabidops
41	658	21.2	503	8	AY092445 Arabidops
42	658	21.2	503	8	AY092446 Arabidops
43	658	21.2	503	8	AY092447 Arabidops
44	658	21.2	503	8	AY092448 Arabidops
45	658	21.2	503	8	AY092449 Arabidops

ALIGNMENTS

RESULT 1